The Role of *CNTNAP2* in the Development & Evolution of the Human Cerebral Cortex

Frances St. George-Hyslop

Leverhulme Centre for Human Evolutionary Studies

Wellcome Trust/CRUK Gurdon Institute

St. Catharine's College, University of Cambridge

September 2020

This dissertation is submitted for the degree of Doctor of Philosophy

Preface

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Frances St George-Hyslop

The Role of *CNTNAP2* in the Development & Evolution of the Human Cerebral Cortex

Frances St. George-Hyslop

The differing cognitive abilities of humans and other primates are accompanied by changes to forebrain neuronal circuit composition and function. This is underpinned by species-specific features of their development. Excitatory neurons in the human cerebral cortex have longer neurites and more elaborate neurite branching than chimpanzees or macaques, which is thought to contribute to differences in neural network performance. To study how differences in expression of single genes contributes to inter-species differences in neuronal form and function, my research focused on Contactin-Associated Protein-like 2 *(CNTNAP2),* a gene important for neuronal differentiation and synapse formation, for which there is accumulating evidence for differential use in the human cerebral cortex compared with other primates. Previous work in mice has found that loss of *Cntnap2* function reduces neurite branching and dendritic spine density *in vitro* and *in vivo*. The potential evolutionary significance of this gene is highlighted by the presence of eight human accelerated regions (HARs), suggesting that there are human-specific aspects to its temporal and spatial expression during cortical development.

The research presented in this thesis used human and non-human primate stem cell-derived forebrain neurons to study several aspects of *CNTNAP2's* function in human cerebral cortex development. We present the first study of neurite outgrowth and neuronal activity in forebrain neurons generated from a human *CNTNAP2* knockout (KO) pluripotent stem cell line. Differentiated human *CNTNAP2* KO neurons have reduced neurite branching relative to wild type cells. Strikingly, the KO neurons were significantly more active - bursting more strongly and more frequently. We also applied a combination of bioinformatic and experimental approaches to show one or more of the *CNTNAP2* HARs may be a gene enhancer.

Our results indicate that loss of function mutations in *CNTNAP2* contribute to human neurodevelopmental diseases through altering neuronal activity. This may be due to changes in neurite branching, and ultimately, to neuronal connectivity. These discoveries suggest a mechanism by which *CNTNAP2* mutations causes disease in affected children, and may have contributed to the evolution of human-specialized brain function. This model will now be invaluable in deciphering the downstream molecular events caused by *CNTNAP2* mutations, and may provide molecular targets for novel treatments in *CNTNAP2* patients.

Acknowledgements

Firstly, I wish to thank my supervisors, Professor Toomas Kivisild and Professor Rick Livesey, for their guidance during my PhD (October 2016 to September 2020). Throughout these four years, Toomas and Rick have been invaluable for my growth as a scientist, and more generally as a person. I cannot thank them enough for their kindness and support, and for giving me the opportunity to test myself as a researcher. I also wish to thank Professor Rob Foley for taking over my supervision at Cambridge, and for his enthusiasm and helpful advice with my project. Similarly, I am grateful to Professor Sadaf Farooqi, Dr Florian Merkle, Dr Neli Atanassova, and Dr Elena Bochukova for mentoring me during my MPhil, and for inspiring me to pursue a PhD.

I owe a great deal of thanks to all members of the Livesey lab: Alessio, Ashley, Ayiba, Christy, Ellie, Elsa, Federica, Francesco, Jayne, James, Lewis, Moritz, Phil, Ravi, Silvia, Steve, and Vickie. I especially thank Alessio, Elsa, and Moritz for all their guidance – without them, the bulk of my research would not have been possible. I am equally indebted to Vickie and Ellie for their help with tissue culture, and to Alessio and Lewis for proof-reading this thesis. I also thank Alex Mörseburg and Sarah Kaewert (Kivisild group) for their assistance with computational genetics.

Outside of the lab, I am grateful to St Catharine's College for being my home away from home. The friends I have made during my time in Cambridge have been my greatest PhD discoveries. I particularly thank Alex Abbott for his support over the last three years, and for his technical help with handling large datasets.

I finally wish to thank my family and friends back in Canada. The never-ending love and encouragement from my parents, Peter and Vera, and my two sisters, Nikki and Sarah, have been vital to the successful completion of my PhD.

Dedicated to my grandparents,

Daphne, Georgina, Noel, and Thomas

my parents,

Peter and Vera

and my sisters,

Nikki and Sarah

Table of Contents

Summary of Figures

Chapter 1

Summary of Figures

Chapter 2

Summary of Figures

Chapter 4

Summary of Figures

Chapter 5

Summary of Figures

Chapter 6

Summary of Figures

Chapter 7

Chapter 8

Summary of Tables

Chapter 1

Chapter 2

Chapter 5

Chapter 6

Summary of Tables

Chapter 7

Abbreviations

ADHD	Attention deficit hyperactivity disorder
ASD	Autism spectrum disorder
Bp	Base pairs
CADD	Combined annotation dependent depletion
CDFE	Cortical dysplasia focal epilepsy
CGE	Caudal ganglionic eminence
CNS[1]	Central nervous system
CNS[2]	Conserved non-coding sequence
CNV	Copy number variation
CRISPR	Clustered regularly interspaced short palindromic repeats
DHS	DNase I hypersensitive site
DIV	Days *in vitro*
gBGC	GC-biased gene conversion
HAR	Human accelerated region
hPSC	Human pluripotent stem cell
ID	Intellectual disability
iPSC	Induced pluripotent stem cell
Kb	Kilo base pairs
KD	Knockdown
KO	Knockout
LGE	Lateral ganglionic eminence
Mb	Mega base pairs
MGE	Medial ganglionic eminence
Mo.	Months
MTG	Middle temporal gyrus
MYA	Million years ago
NPC	Neural progenitor cell
PAM	Protospacer adjacent motif
PFC	Pre-frontal cortex
PTHS	Pitt-Hopkins syndrome
PV	Parvalbumin

RG	Radial glia
RNP	Ribonucleoprotein
scRNA-Seq	Single cell RNA sequencing
snRNA-Seq	Single nucleus RNA sequencing
SLI	Specific language impairment
SNV	Single nucleotide variant
ssODN	Single-stranded donor oligonucleotide
TF	Transcription factor
UTR	Untranslated region

Summary of Appendices

Chapter 1

Introduction

1.1 Introduction to the field of human brain evolution

The cerebral cortex is the principal brain region controlling higher cognition in humans. Although other mammals also possess a cortex, the capacity for specialized functions - including complex language and abstract thought - appear to be unique to *Homo sapiens*. For many years, research into human cortical evolution has been limited. This is partly attributable to the functional and structural complexity of the cortex, but also to the overwhelming size of the primate genome (approximately 3 billion nucleotides). Access to human and primate brain tissue is also tightly restricted for ethical reasons. Over the past decade, however, these roadblocks have been substantially diminished by innovations in two key scientific areas: 1) stem cell neurobiology and 2) comparative genomics.

Identifying the genetic changes underlying human cortical evolution involves the paradigm that changes to the human genome – whether it be changes to single

nucleotides, insertions-deletions, or larger-scale structural changes to chromosomes – have modified the developmental programme of the cortex (1, 2). Because many human-specific features of the cortex first emerge during embryogenesis, differences in development may underpin human cognition. The discovery of methods to generate cortical neurons from pluripotent stem cells ('PSCs'), have revolutionized our ability to study such processes (3, 4). Human and primate PSC-derived cortical neurons can be produced *en masse* to model species differences in brain development, and crucially, to experimentally study these differences in biologically relevant cell types. Phenotypes of interest can be followed over developmental time, while the effects of candidate genes can be robustly tested in loss-of-function (LOF), gain-of-function (GOF), and gene interaction (GI) experiments (5).

Secondly, advances in genomics have enhanced our understanding of DNA function, and highlighted regions of the genome that differ between humans and other primates. Such studies have found that the rate of change in cortical gene expression has accelerated on the human lineage (6). Thousands of genes have been identified that show species-specific expression profiles in the human cortex (6-8). Thousands more have undergone positive selection in humans (9-14), or been duplicated or lost in our species alone (15-21). Despite these breakthroughs, the molecular mechanisms of human cortical evolution are still poorly understood. Especially lacking are functional experiments that make sense of the enormous amount of generated genomic data. Our current challenge is to connect highlighted genes to specific phenotypes, and ultimately, to specific processes of evolution and development.

The task of sifting through billions of base pairs, and determining which sites have led to critical changes in cortical function, is complicated by the fact that most of the human genome still has no understood function. Many comparative studies of human and non-human primate genomes have focused on protein-coding regions, as they are better characterized and simpler to study experimentally (22). However, less than 1% of protein-coding sequences differ between humans and chimpanzees, our closest living relatives (23) (**Figure 1.01**). Over a quarter of all protein sequences are

100% identical, and the remainder differ only in one or two amino acids (many of which are presumed to be functionally neutral) (24). These findings have supported the idea that changes to protein-coding sequences are too rare to fully explain the phenotypic differences between humans and other primates (25). Instead, changes to non-coding DNA are thought to have played a larger role.

When differences in gene expression between humans and other primates are compared in the pre-frontal cortex alone, the most pronounced changes are in genes related to synaptic function (26). Surprisingly, little experimental work has been done to interrogate the role of neural connectivity in human brain evolution. This is particularly curious given the striking morphological and developmental differences between cortical connectivity in humans and other primates (discussed in **Section 1.4**) (27-29). Instead, most studies have focused on the dramatic enlargement of the human frontal cortex (30). However, humans have neither the largest brains nor the most neurons amongst mammals (1). The relationship between brain size and cognitive capacity is also relatively weak (31). Higher brain function, therefore, probably derives from multiple relatively subtle changes to the brain - and not a single major change to one feature. These may include the expansion of specific cortical regions, but also changes to the function and complexity of neural circuits.

Multiple lines of evidence suggest abnormalities in cortical connectivity play a role in cognitive disorders (32) (see **Section 1.3**). These diseases – including intellectual disability (ID), autism spectrum disorder (ASD), and schizophrenia – cause significant distress to individuals and their families. They are also believed to predominantly, if not exclusively, affect humans and target human-specialized behaviors (22). Studies have shown that alterations in dendritic spine density, dendrite morphology, and short and/or long-range neural connections may underlie these disorders. In fact, dendritic spine abnormalities are considered one of the most reliable neuroanatomical correlates of intellectual disability syndromes (32). These associations are unsurprising given that the tightly controlled regulation of synapses is thought to be the basis of learning and memory (33).

Understanding the mechanisms behind human brain evolution will shed light on the longstanding question of what makes us human. As stated, however, it may also inform on the causes and potential treatments of diseases affecting the human cortex (22). In light of these prospects, I set out to investigate the role of cortical connectivity in both human brain evolution and disease. My thesis focuses on the role of one gene, **contactin-associated protein-like 2** ***(CNTNAP2)***, which stood out as an understudied candidate worthy of further examination. *CNTNAP2* encodes CASPR2, a member of the neurexin family of type I transmembrane proteins. A full introduction to *CNTNAP2* is given in **Chapter 2**, however, in brief *CNTNAP2* was prioritized for the following key reasons:

i) *CNTNAP2* has been shown to increase dendritic branching and spine density *in vitro* and *in vivo* (34-36);
ii) It has higher expression levels in the frontal cortex of humans compared to non-human primates (37-39);
iii) It contains signatures of positive selection within human populations (40, 41);
iv) It has seven 'Human Accelerated Regions' (HARs), sequences that are highly conserved across most vertebrates, but differ dramatically in humans (9, 42);
v) It is mutated in disorders affecting cognition (e.g. intellectual disability and autism spectrum disorder) (43, 44).

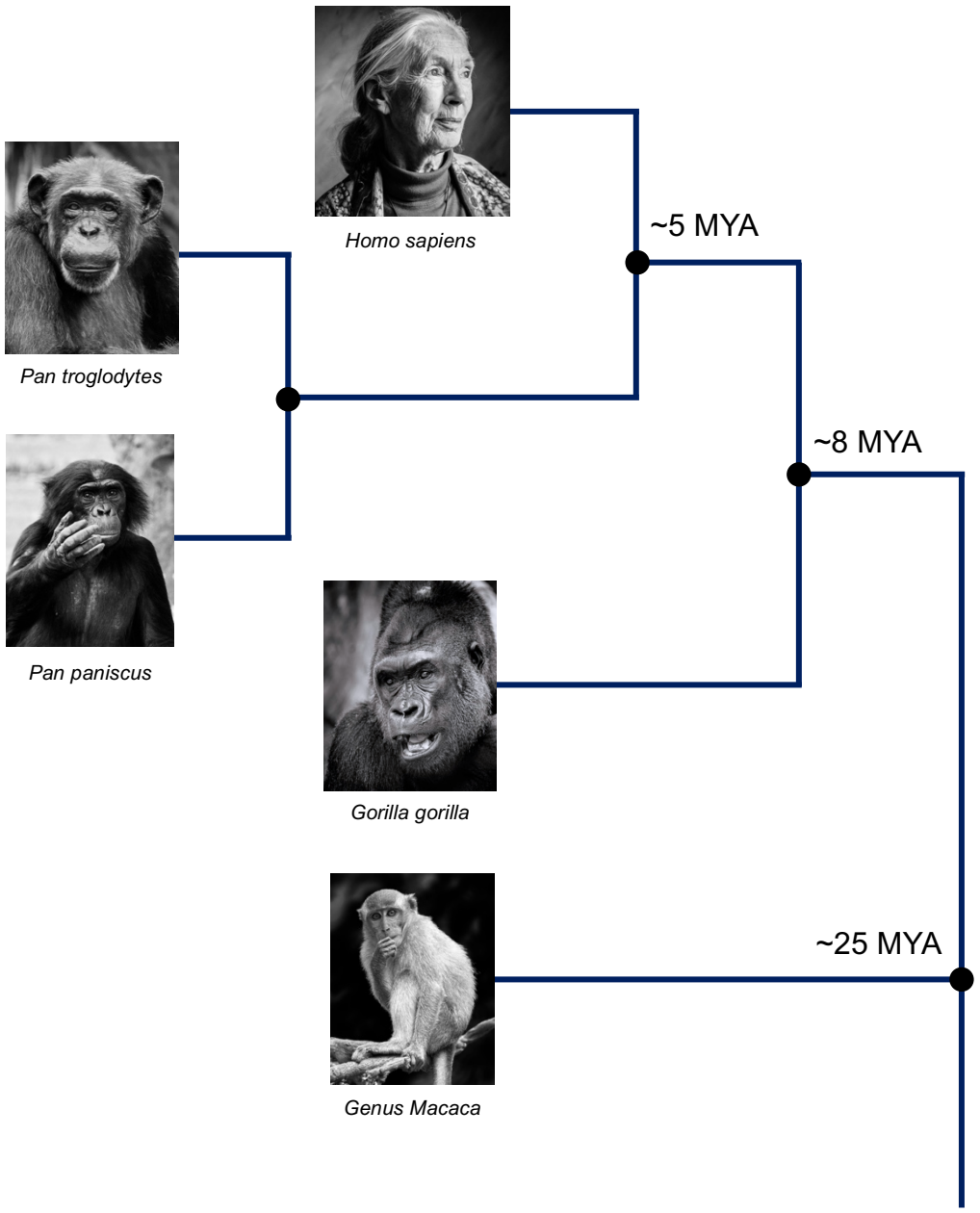


Figure 1.01. Phylogeny of extant primates

Phylogenetic relationships between humans and a selection of our closest living primate relatives. Our closest relatives are chimpanzees *(Pan troglodytes)* and bonobos *(Pan paniscus*). These species share ~99% genome similarity with *Homo sapiens*. The dates shown represent the time to the most recent common ancestor between lineages. MYA: millions of years ago.

1.2 A brief evolutionary history of *Homo sapiens*

Figure 1.01 shows the relationships between humans and our living primate relatives. Our closest relatives are the chimpanzee *(Pan troglodytes)* and the bonobo *(Pan paniscus)*. We share 98.8% genome identity with both species, and diverged apart ~5 million years ago (MYA) (24, 45). Divergence dates are based on a combination of anatomical, fossil, and molecular methods (46). Following on from *Pan* is our next closest kin, the gorilla *(Gorilla gorilla)*. Humans and gorillas diverged slightly earlier, approximately 8 MYA, and are less genetically similar (~96% genome identity) (45, 47). Continuing down the phylogenetic tree are the orangutans *(Pongo pygmaeus)* and the gibbons (genus *Hylobates)*. These species are our fourth and fifth closest relatives respectively, having split ~15 and ~21 MYA (6, 45). Finally, among the more distantly related Old World Monkeys *(Cercopithecoidea)*, are the macaques (genus *Macaca*). Macaques are noteworthy for their frequent inclusion in biomedical research (48). Macaques and humans split 25-28 MYA, and have even less genomic similarity, with ~91% identity (49). Differences between the brains of humans and living primates will be discussed at depth in **Section 1.4**.

The term **hominin** refers the sub-family or tribe (depending upon the precise classification used) of species that are more closely related to modern humans than to other living great apes (50). Hominin origins is therefore usually placed at the human-chimpanzee split, some 5-8 MYA (51, 52). Precisely which species is the first member of the clade (of the fossils identified) is still debated. The lack of consensus is caused by differing inclusion criteria and by the incomplete nature of the fossil record (for a full review, see Boyle & Wood (50)). Three taxa are frequently named as candidates for the first known hominin: 1) ***Sahelanthropus tchadensis***, 2) ***Orrorin tugenensis***, and 3) ***Ardipithecus, A. kadabba*** and ***A. ramidus*** (see **Figure 1.02**) (50). These species lived in Africa between 7 - 4 MYA, and have similar anatomical features to other hominins (46). Each provides some, but incomplete, evidence for hominin traits. For example, *S.*

tchadensis had an anteriorly placed foramen magnum[1] (suggestive of bipedalism), while *O. tugenensis* had human-like femur bones, and *A. ramidus* had reduced canine teeth (46). However, these species also had small endocranial volumes. *S. tchadensis* and *A. ramidus* ranged around 300-365 cm^3, making their brains similar in size to present-day chimpanzees (53).

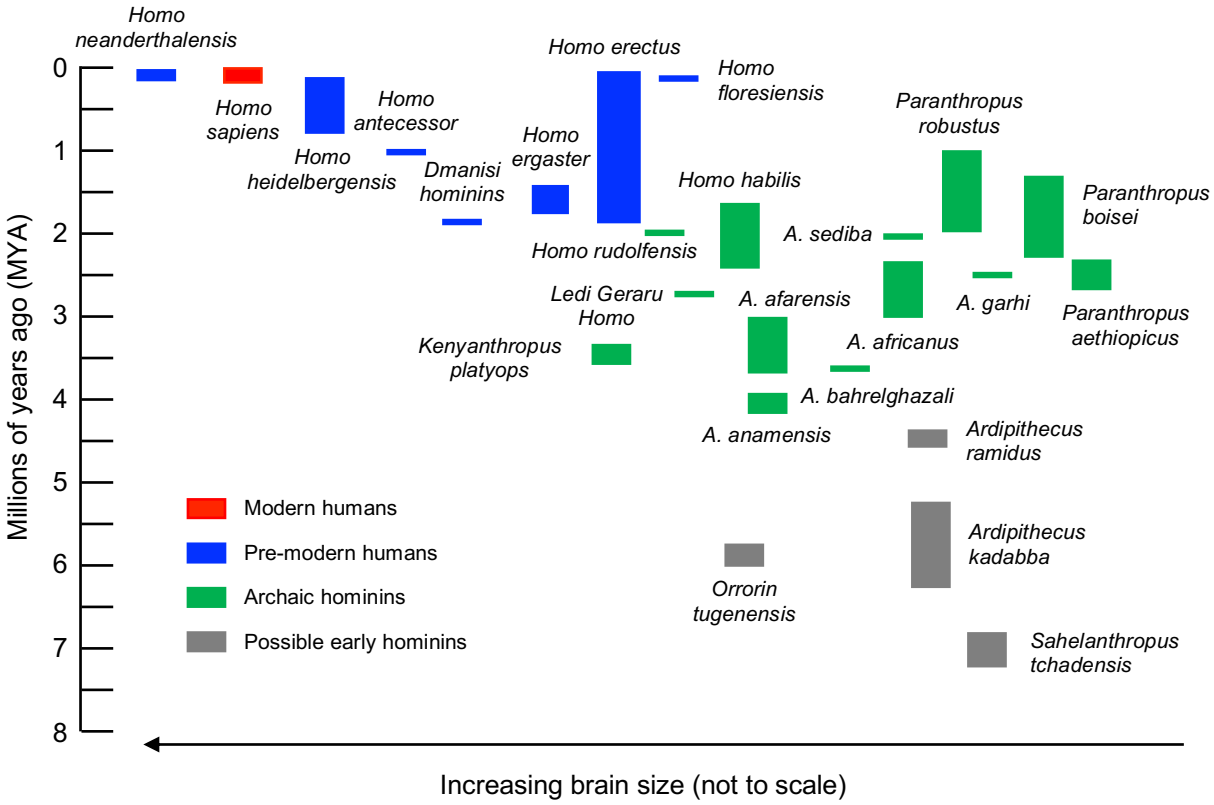


Figure 1.02. Timescale of hominin evolution

Hominins are species more closely related to modern humans than other living great apes. Hominins are plotted by their estimated date of existence (y-axis) and estimated brain size (x-axis). Taxa with larger brains are shown on the left. This plot recognizes a larger rather than a smaller number of species. The term 'possible hominins' refers to species for which there is ongoing debate about their inclusion in the hominin clade. Figure adapted from Boyle & Wood (2017).

[1] Foramen magnum: the opening at the base of the skull where the spinal cord passes.

The first definite hominins were ***Australopithecines,*** notably ***A. anamensis*** at 4.2 MYA (50). This and related species had relatively human-like dental anatomy, and a post-cranial anatomy that suggests bipedalism in combination with arboreal activity (54). The *Australopithecus* radiated into a number of forms in East and South Africa, and persisted until around 2.0 MYA. From about 2.6 MYA the more robust and dentally specialised ***Paranthropus*** evolved and radiated, surviving until around 1.0 MYA. *Paranthropus* had a slightly larger brain than *Australopithecus* - with an endocranial volume between 400-650 cm^3 as compared to ̓350–550 cm^3 (50). Both lineages were of a similar size, with a body mass between 25-60 kg.

The genus *'Homo'* first appeared somewhere between 3 - 2 MYA (55). ***Homo habilis*** provides the most convincing evidence for early *Homo* (54)*. H. habilis* had a larger braincase than *Australopithecus* or *Paranthropus* (volume estimates range from 450–800 cm^3), but still possessed ape-like traits, including long arms and prognathic faces (50). *H. habilis* is also the first hominin to use stone tools persistently, which is the source of their name 'handy man'.

The early forms of *Homo* are clear transitional forms from *Australopithecus*, but by shortly after 2 MYA, a species with much more human features evolved. This is species known in Africa as ***Homo ergaster*** by many, but is also included more broadly into the widespread taxon ***Homo erectus***. From around 1.8 MYA hominins dispersed out of Africa and into the warmer parts of Eurasia. *H. erectus* had a much more human-like body shape and proportions. This species would have been bipedal in ways almost identical to modern humans. Their technology was also more complex and persistent. It was with *H. erectus* that endocranial volume increased significantly in the hominin lineage, with volumes ranging between 650–1300 cm^3 (50, 56).

Once widespread across Africa and Eurasia, hominins probably diversified into different species and sub-species, but the evidence for this is relatively sparse. It appears that in Africa a new lineage, ***H. heidelbergensis***, evolved around 700 KYA, and also dispersed out of Africa and into Eurasia (notably, to much higher latitudes). *H.*

heidelbergensis had larger bodies and brains than previous hominins – their average mass was 67 kg, while their endocranial volume ranged from 1000–1250 cm^3 (50). They are also known for their ability to build shelters (57) and control fire (58).

H. heidelbergensis is thought to be the common ancestor of three major lineages that evolved and dominated for the last 500 KYA: the **Neanderthals** and **Denisovans** in Eurasia, and **modern humans** in Africa. Neanderthals and Denisovans are our closest relatives, sharing 99.7% of our genome (59). Neanderthal brains were larger than modern humans (1100–1750 cm^3), and proportional to their brawnier bodies (65-85 kg). Their cognitive abilities were also sophisticated; they may have worn clothing, buried their dead, and created art and/or music (60). No earlier hominins are known to have practiced such behavior. Regarding Denisovans, little information is known about them – only seven skeletal fragments have been discovered (61). Ancient DNA (aDNA) sequencing revealed they were a distinct, but closely related species, to Neanderthals. Genetic evidence suggests there was interbreeding between modern humans, Denisovans, and Neanderthals. DNA transferred from Denisovan/Neanderthal into humans ('introgressed') may even have conferred adaptive advantages in some populations (61).

Our own species, *Homo sapiens*, appeared in Africa some 200 KYA – possibly also from *Homo heidelbergensis* (57). While all other hominins eventually went extinct (*H. erectus* around 30 KYA and *H. neanderthalensis* around 40 KYA), *Homo sapiens* survived and expanded (50). By 100-40 KYA, our ancestors migrated out of Africa and into Eurasia, Australia, and the Americas (62). Current evidence suggests there were multiple **'Out of Africa'** migrations, each involving multiple small **bottlenecks** (a rapid reduction in genetic diversity) (63). Of all human populations, Africans have the greatest genetic diversity. This idea has also led to the **serial founder model**, whereby human populations that migrated out of Africa (and contained only a fraction of the total genetic diversity) seed into sub-populations, which in turn seed into additional sub-populations – each containing a smaller and smaller proportion of

genetic variation (63). Such changes to the gene pool have likely had important consequences for human evolution.

Lastly, precisely when and how humans acquired their cognitive powers is still unclear. Two models have been proposed – one that argues this was a gradual process and another that argues it was rapid (31). The latter theory, called the **human revolution model** (64), is based on the sudden and nearly simultaneous explosion of artistic, technological, and cultural artefacts from around 50-40 KYA (e.g. musical instruments, art, specialized hunting, etc.). However, the more recent discoveries of similar findings from 100-75 KYA (i.e. earlier in time), has shaken the credibility of this model (65). Instead, a slow and gradual acquirement of higher brain function is the currently favored theory.

Since soft tissues, behavioural expressions and spoken language do not preserve directly, our knowledge of hominin brains must be extrapolated from other sources. These sources include skeletal remains, material remains (e.g. artifacts), and ancient DNA (66). Innovations in anthropology, archaeology, and paleogenomics have revolutionized our perceptions of archaic humans. However, without access to real brain samples our ability to understand hominin brain function remains limited. This challenge is compounded by our inability to observe hominin behavior. For these and other reasons, my thesis focuses on comparisons between humans and living primates.

1.3 Structure and development of the human cerebral cortex

1.3.1 Overview of the human cerebral cortex

The cerebral cortex is the convoluted outer covering of the cerebrum. There are an estimated 16 billion neurons in the human cortex, which form 160 trillion synapses between them (1). Cortical neurons can be classified into two major types: 1) excitatory neurons (which produce the neurotransmitter **glutamate**) and 2) inhibitory interneurons (which produce gamma-aminobutyric acid, or **GABA**) (67). Approximately 80% of all cortical neurons are glutamatergic, with the remaining 20% GABAergic (68).

The cortex is divided into six layers or **'laminae'** (see **Figure 1.03**). Each layer contains sub-populations of neurons with unique morphologies, connections, and patterns of gene expression. Neurons form local 'microcircuits' between cortical layers, as well as more distant intra-cortical and extra-cortical connections (69, 70). Such layering is found in the cortex of all mammals – although the size and organization of the cortex differs greatly between species (see **Section 1.4**) (71). In addition to its laminar (horizontal) arrangement, the cortex is further organized into vertical units called **'mini-columns'** (69). Mini-columns contain ~80-100 cells that are heavily inter-connected and clonally related. These columns are hypothesized to be the smallest functional units capable of information processing.

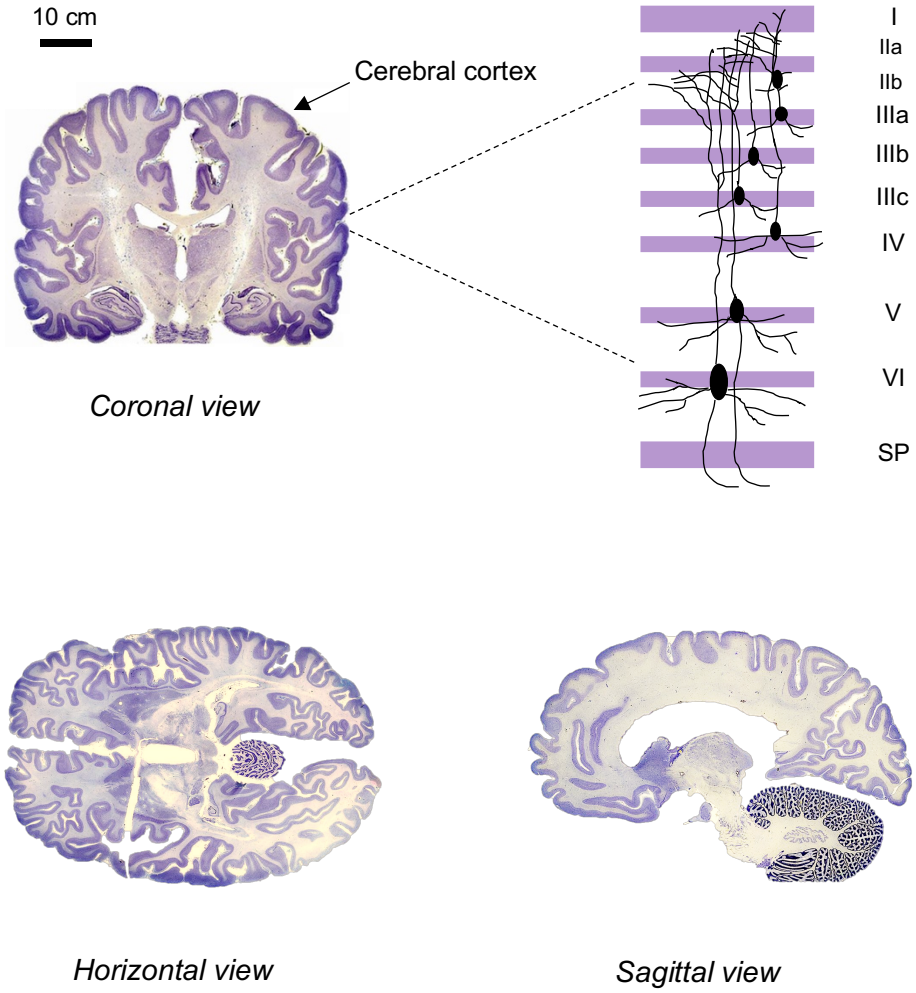


Figure 1.03. Structure of the primate cerebral cortex

The cerebral cortex is the convoluted outering covering of the brain (shown here in purple). It is responsible for higher cognitive functions including complex language and abstract thought. The primate cortex is composed of six layers of neurons that form intra- and extra-cortical connections. There have been marked changes to the size, structure, and organization of the cortex on the human lineage. These changes will be discussed in detail in **Section 1.4**. Sections taken from the University of Wisconsin Comparative Mammalian Brain Collection. Illustration of cortical layers is adapted from Hill & Walsh (2005). SP: sub-plate.

1.3.2 Structure of the human telencephalon

By embryonic week 6, the human brain begins to develop at the anterior neural tube (72). The brain takes it form from three major vesicles: 1) the **forebrain** (*prosencephalon*), 2) the **midbrain** (*mesencephalon*), and 3) the **hindbrain** (*rhombencephalon*) (73) (**Figure 1.04a**). The forebrain is further divided into the **telencephalon** (which produces the cerebral cortex) and the **diencephalon** (which generates the thalamus, hypothalamus, and pituitary gland). Each of the major brain divisions are also associated with particular genes/patterning factors that define them (examples listed in **Figure 1.04**).

Within the telencephalon, transcription factors pattern three broad domains: 1) the dorsal telencephalon (also called the **'dorsal pallium'**), 2) the medial telencephalon (**'medial pallium'**), and 3) the ventral telencephalon (**'sub-pallium'**) (**Figure 1.04b**). These regions will diversify into the many structures of the adult forebrain. The dorsal telencephalon is the site of the future cerebral cortex – it generates neural progenitor cells and differentiated excitatory neurons. The medial pallium contains the primordial hippocampus, the cortical hem (a transient signalling centre), and the choroid plexus (a secretory tissue that generates cerebral spinal fluid) (74). Lastly, the basal ganglia will form from the sub-pallium, which is separated into three **'ganglionic eminences'**: the medial ganglionic eminence (MGE), the lateral ganglionic eminence (LGE), and the caudal ganglionic eminence (CGE).

While excitatory (glutamatergic) cortical neurons form in the dorsal pallium, inhibitory (GABAergic) cortical neurons form separately in the sub-pallium (75). Each of the ganglionic eminences is responsible for producing particular sub-types of interneurons. For example, the MGE generates **parvalbumin (PVB)** interneurons and the CGE generates **calretinin (CALB2)** interneurons (72). Once differentiated, these cells migrate dorsally to the forming cerebral cortex.

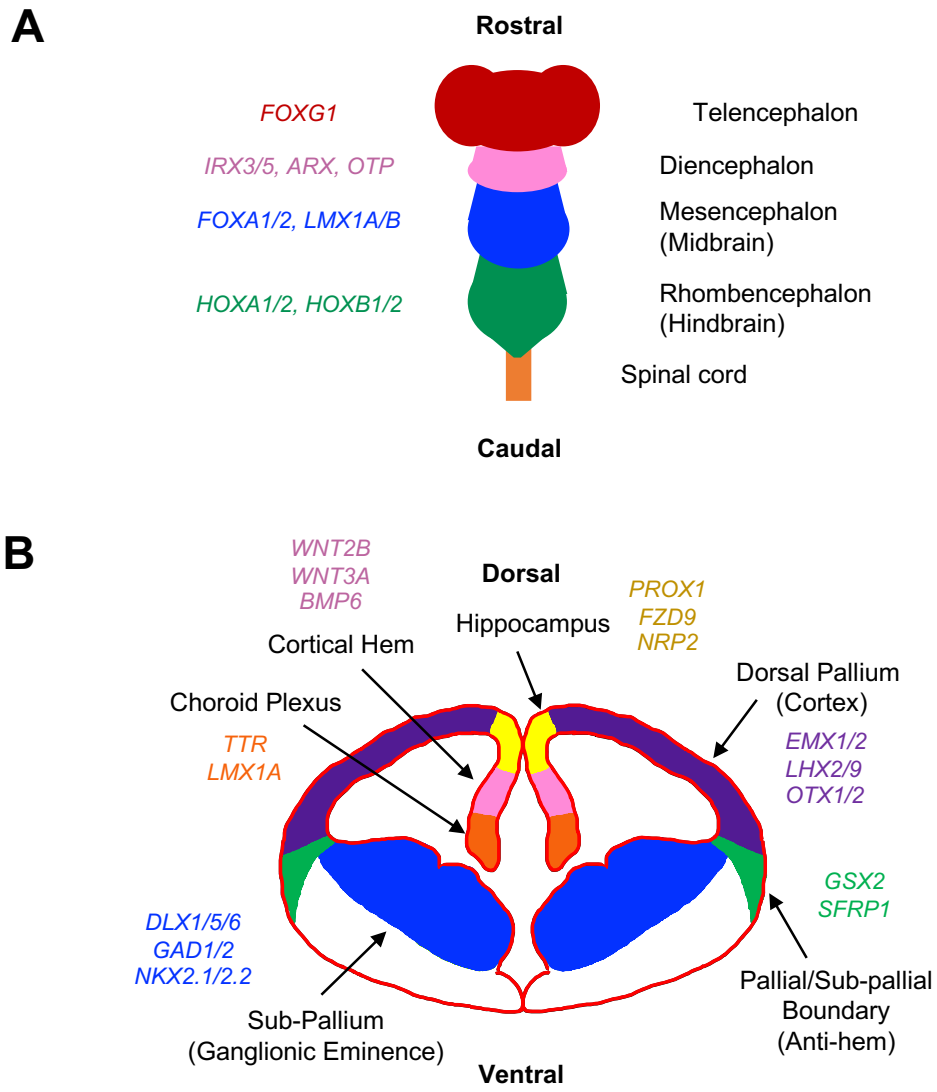


Figure 1.04. Structures of the developing brain

(A) Schematic of the early embryonic brain. The forebrain consists of both the telencephalon and the diencephalon. The telencephalon gives rise to the cerebral cortex, while the diencephalon produces the hypothalamus, thalamus, and pituitary gland. The midbrain produces the tectum and tegmentum, and the hindbrain generates the cerebellum and brainstem. Key genes associated with each structure are shown on the left. (B) Frontal slice of a human embryonic brain (embryonic week 11.5). The dorsal pallium gives rise to the cortex, and the medial pallium to the hippocampus and choroid plexus. The sub-pallium consists of three ganglionic eminences (not shown): lateral, medial, and caudal. Cortical interneurons are produced here and migrate to the dorsal pallium during development. Major genes associated with each region are shown in corresponding colours.

1.3.2 Overview of human corticogenesis

The cerebral cortex originates from a single layer of **neuroepithelial stem cells (NSCs)** (76) (**Figure 1.05a**). Neurogenesis begins when NSCs switch from symmetric (proliferative) division to undergo asymmetric (differentiative) division. The terms 'symmetric' and 'asymmetric' refer to the fates of the resulting two daughter cells. Before neurogenesis, NSCs divide symmetrically to form two additional NSCs. This increases the surface area and thickness of the forebrain primordium. In asymmetric division, one or both of the daughter cells differentiate into a distinct cell type. The first daughter cell may remain an NSC or transition into a **radial glia (RG)**(77). In contrast, the second daughter cell will either become an **intermediate progenitor (IP)** or a **post-mitotic cortical neuron**. All neurons in the cortex are post-mitotic, meaning they will not divide again during their lifetimes. **Table 1.01** summarizes key genes that are associated with each cell type.

NSCs, RG, and IPs are all classified as **neural progenitor cells (NPCs)**. Like NSCs, RG can self-renew via proliferative division or become IPs/excitatory neurons via differentiative division. IPs can also self-renew, but are only capable of differentiating to neurons (i.e. and not to RG) (75). Neural progenitors reside in separate anatomical locations within the dorsal pallium. NSCs and RG occupy the **ventricular zone (VZ)**, while IPs populate the **sub-ventricular zone (SVZ)** (**Figure 1.05b**). Confusingly, the SVZ is located dorsal to (i.e. above) the VZ, closer to the **pial surface** (top of the cortex) and further from the **ventricular surface** (bottom) (78).

Before the formation of the SVZ, however, the VZ is first invaded by **'Cajal-Retzius (CR) neurons'** (75) (**Figure 1.05b**). CR neurons secrete the extracellular glycoprotein, **reelin**, which plays an important role in cell migration and cortical lamination (79). Specifically, in reelin deficient mice ('reeler' mice) the six-layered structure of the cortex is inverted. Cajal-Retzius neurons are mainly generated outside the cortex - in the hem, the choroid plexus, and the pallial/sub-pallial boundary (also

known as the **anti-hem**) (80). Once they arrive at the dorsal pallium, they settle on top of the VZ to form the **pre-plate**.

The first *de facto* cortical neurons accumulate under the pre-plate (now called the **marginal zone (MZ)**) by human embryonic day 50 (E50) (75). RG extend long **'radial processes'** from the VZ directly to the pial surface (**Figure 1.05c**). These processes act as a scaffold, guiding migrating neurons to their destination in the **cortical plate**. Neurons in the marginal zone will become **cortical layer I**, while the cortical plate becomes **cortical layers II-VI** (72). Located directly under the cortical plate is the **sub-plate**, which forms the deepest cortical layer, **layer VIb**. Crucially, layer VI neurons are generated first, followed by layer V, until the last formed neurons occupy layer II. Since layer II is located closer to the pial surface, this means superficial layers must migrate past older, deeper layers. The majority of neuronal migration occurs between the third and fifth months of human gestation (75). By the seventh gestational month, each of the six cortical layers are clearly defined.

After the completion of neurogenesis, RG begin to produce glial cells (e.g. oligodendrocytes, microglia, and astrocytes). Human **gliogenesis** begins at gestational week 25 and ceases by early childhood (81). Glial maturation, however, continues for a much longer period of time. For example, the myelination of axons by oligodendrocytes can take up to 20 years to complete (72).

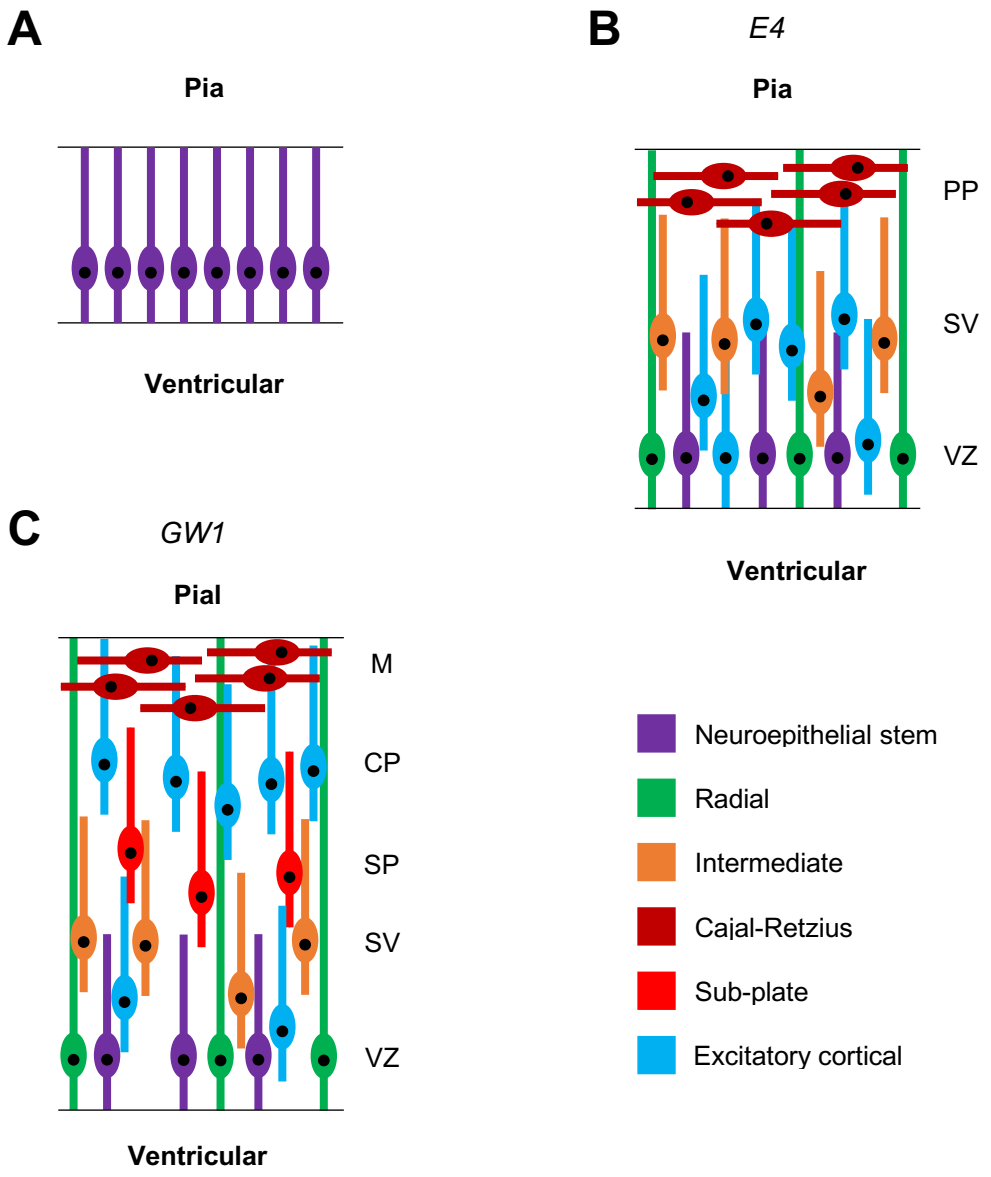


Figure 1.05. Transitional phases of the developing human cortex

(A) The cerebral cortex forms from a single layer of neuroepithelial stem cells. (B) By embryonic day 40 (E40) the neuroepithelium has divided into three germinal layers: the ventricular zone (VZ), the sub-ventricular zone (SVZ), and the pre-plate (PP). The VZ contains neural stem cells and radial glia. The SVZ contains intermediate progenitors, and the pre-plate contains Cajal-Retzius neurons that have migrated from outside the cortex. Radial glia extend long processes from the ventricular surface to the pial surface. Neurons use these processes to migrate dorsally and settle in the cortical plate (CP). (C) By gestational week 14 (GW14) the cortical plate and the sub-plate (SP) have formed (and the PP is now called the marginal zone (MZ)). By the seventh month, the cortical plate is clearly divided into its canonical six layers (not shown). Figures adapted from Götz (2001).

Cell type	*Associated genes*	*Symbol*
Neuroepithelial stem cell (NSC)	E-cadherin	*CDH1*
	Hairy and enhancer of split-1 or -3	*HES1, HES3*
	Nestin	*NES*
	Notch receptor 1	*NOTCH1*
	Occludin	*OCLN*
	SRY-Box transcription factor 2	*SOX2*
Radial glia (RG)	Fatty acid binding protein 7	*FABP7*
	Glial fibrillary acidic protein	*GFAP*
	Hairy and enhancer of split-1 or -5	*HES1, HES5*
	Nestin	*NES*
	Paired box 6	*PAX6*
	SRY-box transcription factor 2	*SOX2*
	Vimentin	*VIM*
Intermediate progenitors (IPs)	Achaete-scute homolog 1	*ASLC1*
	Eomesodermin	*EOMES*
Cajal-Retzius neurons (CR neurons)	Reelin	*RELN*
	Tumour protein P73	*TP73*
Deep layer cortical neurons	B cell leukemia 11b	*BCL11B*
	Fez family zinc finger protein 2	*FEZF2*
	Forkhead box protein P2	*FOXP2*
	SRY-box transcription factor 5	*SOX5*
	T-box brain transcription factor 1	*TBR1*
Upper layer cortical neurons	cut like homeobox 1 or 2	*CUX1, CUX2*
	POU class 3 homeobox 2	*POU3F2*
	Special AT-rich sequence binding protein 2	*SATB2*

Table 1.01. Key genes expressed by cortex-associated cell types

These genes are commonly used to identify particular cell types in the cortex. Although they are frequently referred to as ‘marker genes’, they serve important biological functions during cortical development. For example, PAX6 is a paired-box transcription factor that is essential for correctly patterning the dorsal telencephalon.

1.3.3 Cortical synaptogenesis

By mid-gestation, cortical neurons have begun to extend axons and dendrites (referred to as **'neurites'** collectively) (82). Neurites develop from membranous projections called **filopodia** and **lamellipodia**. At the tips of these structures are **'growth cones'**, which guide the filopodia/lamellipodia towards their target. Outgrowth is aided by **guide-post cells** (immature neurons or glia) that secrete molecules to attract or repel growth cones to their destinations (81). Classical guidance molecules include members of the ephrin, netrin, semaphorin, and slit families (82). Once at their target cells, neurites form connections called **synapses**.

In its simplest form, a synapse consists of a **pre-synaptic** axon (where neurotransmitters are synthesized and emitted), a **synaptic cleft** (where neurotransmitters are secreted into), and a **post-synaptic** dendrite (which receives and responds to the signal) (83) (**Figure 1.06**). When excitatory neurotransmitters are received, they depolarize the post-synaptic cell to trigger downstream neuronal activity. Inhibitory neurotransmitters do the opposite – they hyperpolarize the cell to reduce neuronal activity. Prenatally, most synapses form between an **axon terminal** and the main **dendritic shaft** (83). This shifts postnatally, where approximately 90% of excitatory synapses are formed on **dendritic spines** (81). Spines are small, actin-rich protrusions (~2-3 μm in length) that extend off dendritic shafts. They contain a **'post-synaptic density'**, which is a sub-cellular region housing neurotransmitter receptors, ion channels, and signaling molecules (84). Spines are thought to increase the surface area for synaptic contacts to be made (85), and to create spatially isolated post-synaptic compartments (86). This would allow the biochemical responses of individual synapses to be separated on the same neuron (i.e. synapses made with the axons of other pre-synaptic neurons). Amongst primates, spine density and dendritic branching are greatest in higher order processing areas like the pre-frontal cortex (28).

The first cortical synapses are formed in the pre-plate around human gestational week 5 (87) (see **Figure 1.07a**). By the third trimester, neurite arborization and synaptogenesis begin to accelerate in the cortical plate (88). However, most prenatal neural connections are transient, and the bulk of synaptogenesis occurs after birth (89). For the first two postnatal years, neurites and synapses are vastly overproduced. The timing of peak synapse density varies by cortical region, but in the frontal cortex occurs between 8 to 15 months (88). From early childhood through to adolescence, synapses will be strengthened or eliminated in an activity-dependent manner (90). The number of synapses that are pruned is substantial; in the adult visual cortex, for example, only ~50% of the synapses present at maximum density (age 4 months) are retained (83). This process – which occurs in all primates - is thought to refine the brain's circuitry, and allow for more efficient cognitive functioning. It also coincides with developmental milestones, including the onset of speech and language or the formation of episodic memories (**Figure 1.07b**) (87).

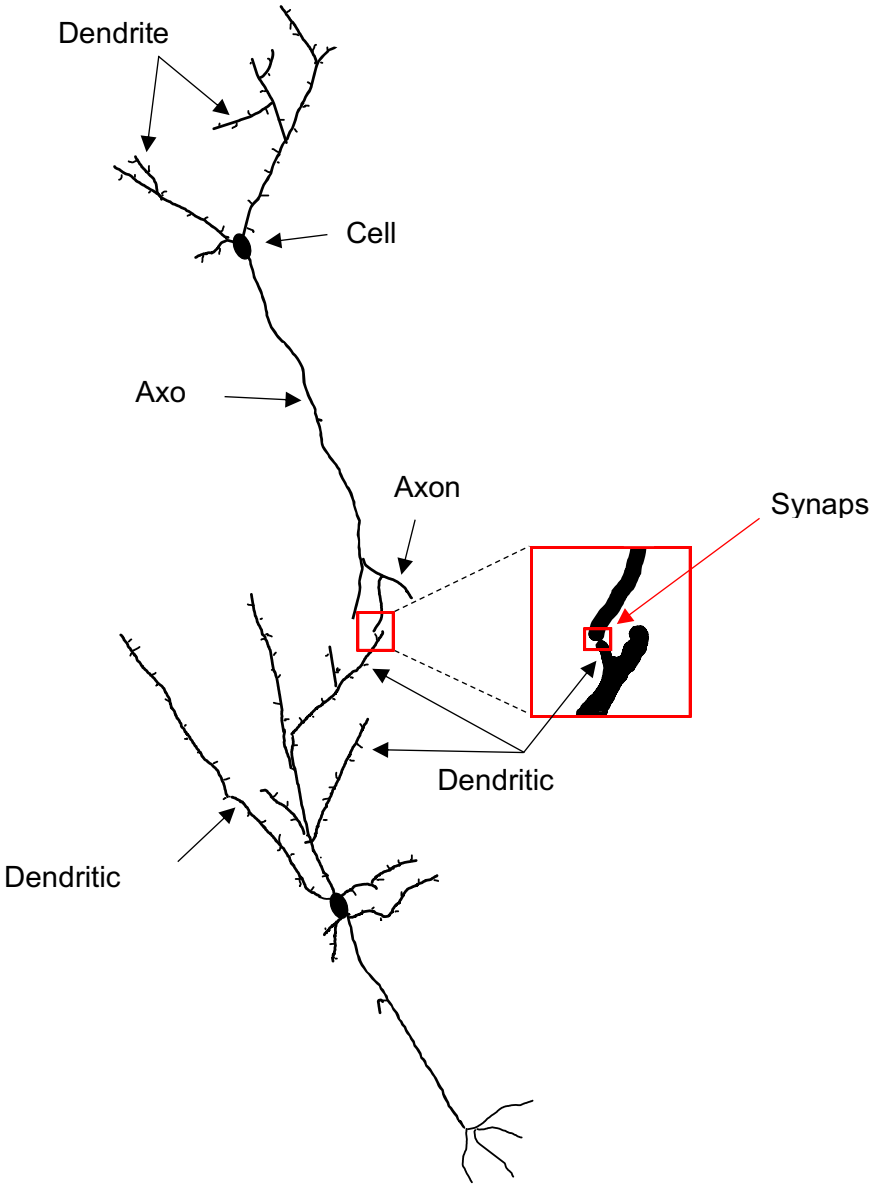


Figure 1.06. Structures involved in excitatory cortical synapses

A synapse consist of a pre-synaptic axon (where neurotransmitters are synthesized and emitted), a synaptic cleft (where neurotransmitters are secreted into), and a post-synaptic dendrite (which receives and responds to the signal). Prenatally, most synapses form between an axon terminal and the main dendritic shaft. This shifts postnatally, where approximately 90% of excitatory synapses are formed on dendritic spines. Spines are small, actin-rich protrusions (~2-3 μm in length) that extend off dendritic shafts. They contain a 'post-synaptic density', which is a sub-cellular region housing neurotransmitter receptors, ion channels, and signaling molecules. Spines are thought to increase the surface area for synaptic contacts to be made, and to spatially isolate post-synaptic signals. Diagram not to scale.

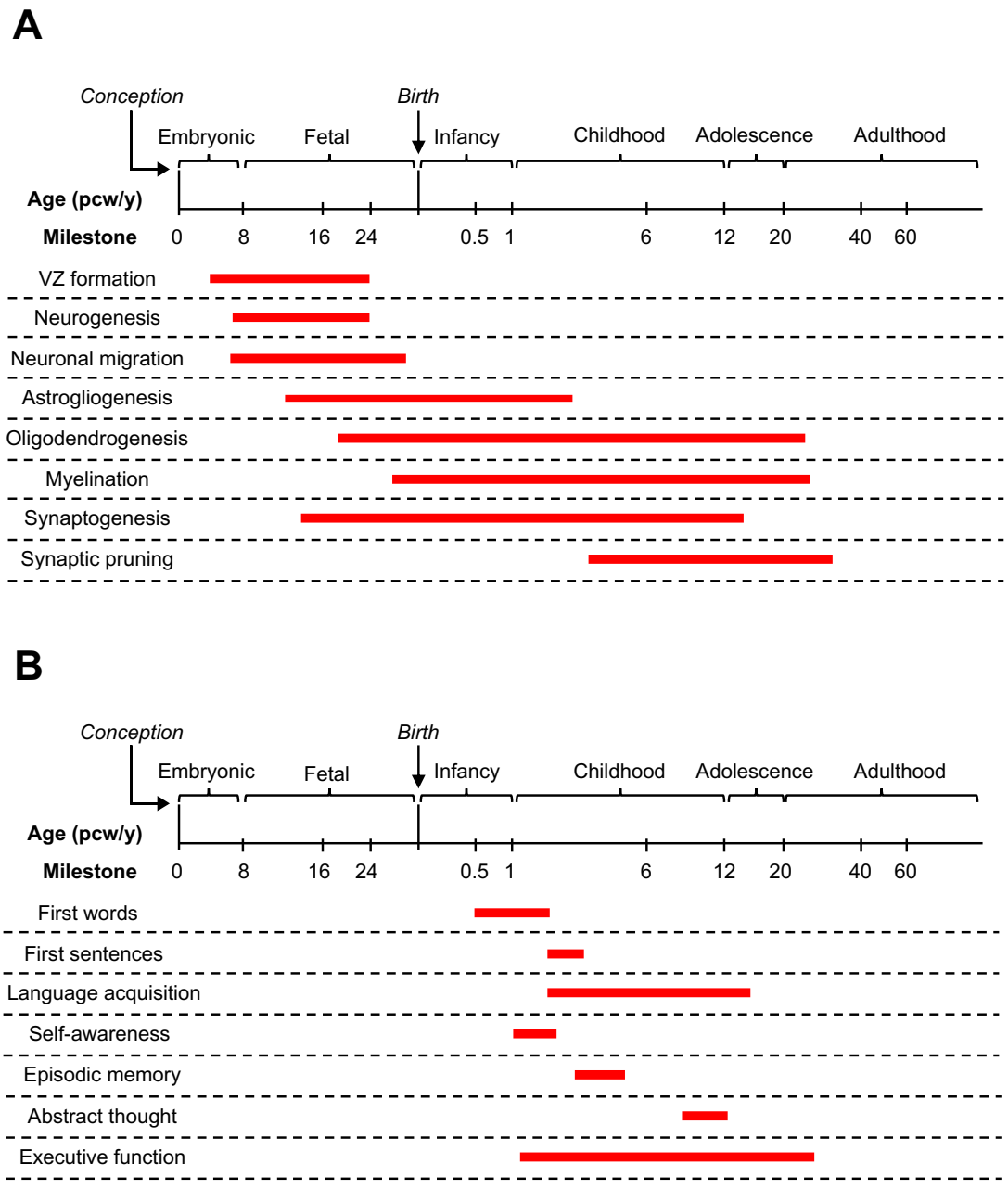


Figure 1.07. Time-course of human cortical development

(A) Approximate timings of major milestones in corticogenesis. Although certain processes continue (on a lesser scale) throughout life, this figure shows the major periods during which the bulk of development occurs for each process. (B) Approximate timings of major milestones in cognitive development. These landmarks overlap with the later stages of corticogenesis, including axon myelination, synaptogenesis, and synapse pruning. PCW: post-conception weeks; Y: years.

1.3.4 Synapse dysfunction and disease

Synaptogenesis is a critical period for human cognitive function. This dependency, however, may have also made our brains vulnerable to disease (91). As mentioned in **Section 1.1**, a number of neurological/psychiatric disorders are thought to involve aberrant synaptic function. These include (but are not limited to) schizophrenia, autism spectrum disorder (ASD), and intellectual disability (ID). Many of these illnesses manifest in childhood or adolescence, overlapping key stages of synapse formation and pruning (**Figure 1.08a**). These disorders are also extremely complex – symptoms can vary largely, and the underlying causes typically involve intricate combinations of environmental and genetic factors (92). That said, etiological theories tend to converge on the idea of an **excitation/inhibition (E/I) imbalance** (93).

For example, schizophrenia is hypothesized to be a disorder of excessive synapse elimination – and therefore 'hypo-connectivity' (94) (**Figures 1.08-1.09** and **Table 1.02**). Symptoms include delusions, hallucinations, disorganized behavior, and blunted emotions (95). It is estimated to affect ~1% of the world's population, and increases the mortality rate of those affected by 1.5-fold (30% of this increase is attributable to suicide) (96). Schizophrenia is typically diagnosed in adolescence or young adulthood, which coincides with the peak period for synapse pruning. The underlying causes of the disease are still an active area of research. However, schizophrenia has been connected to abnormal patterns of brain connectivity, particularly between the prefrontal cortex and the striatum, thalamus, and limbic system (94, 97). Studies of patient brain samples have also found a reduction in the density of dendritic spines in the pre-frontal cortex (98-101). Similar reductions in dendritic branching (102), and in the expression of genes associated with synaptic function (particularly glutamate signalling), have also been identified (103, 104). These findings could explain the reduced cortical volume observed in schizophrenic patients (94). Since these reductions occur without significant changes to neuronal

number, this suggests that loss of axons/dendrites may be accounting for these phenotypes.

ASD is a second neurodevelopmental disorder associated with an E/I imbalance. Symptoms of autism include repetitive or restricted behaviors, deficits in communication, and abnormal socialization (105). ASD frequently occurs alongside other neurological disorders, including intellectual disability and epilepsy (106). It has a similar prevalence to schizophrenia, affecting 0.5-1% of the global population, but with a greater tendency to affect males (105). Cortical tissue from autistic donors has revealed significantly increased dendritic spine density and dendritic branching (107, 108). As with schizophrenia, many genes associated with the disorder are involved in synaptic transmission (**Table 1.03**) (109). In fact, of the 102 most significant rare variants associated with ASD to date (110), 24 are involved in neuronal communication (e.g. *SYNGAP1*, a post-synaptic protein, and *SCN2A*, a voltage-gated sodium channel). ASD also presents early in childhood when synapse growth is at its maximum. Autism could therefore be caused by either excessive synaptogenesis or by insufficient pruning. The true picture, however, is probably more complex than global hyper-excitability. The brains of ASD patients show local areas of over-excitation, but this is balanced by underconnectivity at longer ranges (92). Moreover, there have been reports of autistic individuals having reduced – not increased – dendritic branching (111).

For the majority of neurodevelopmental disorders, further research is needed to clarify causal mechanisms and unearth potential therapies. While post-mortem tissue and animal models are important to model complex, *in vivo* systems, the use of PSC-derived cortical cells from patient and control lines has become increasingly useful (94, 112, 113). Instead of a 'snapshot' view of deceased tissue, for example, these *in vitro* systems provide a means to investigate disease processes in a dynamic, living, and developing cell line. The prospects of human PSC-derived models will be discussed further in **Section 1.6**.

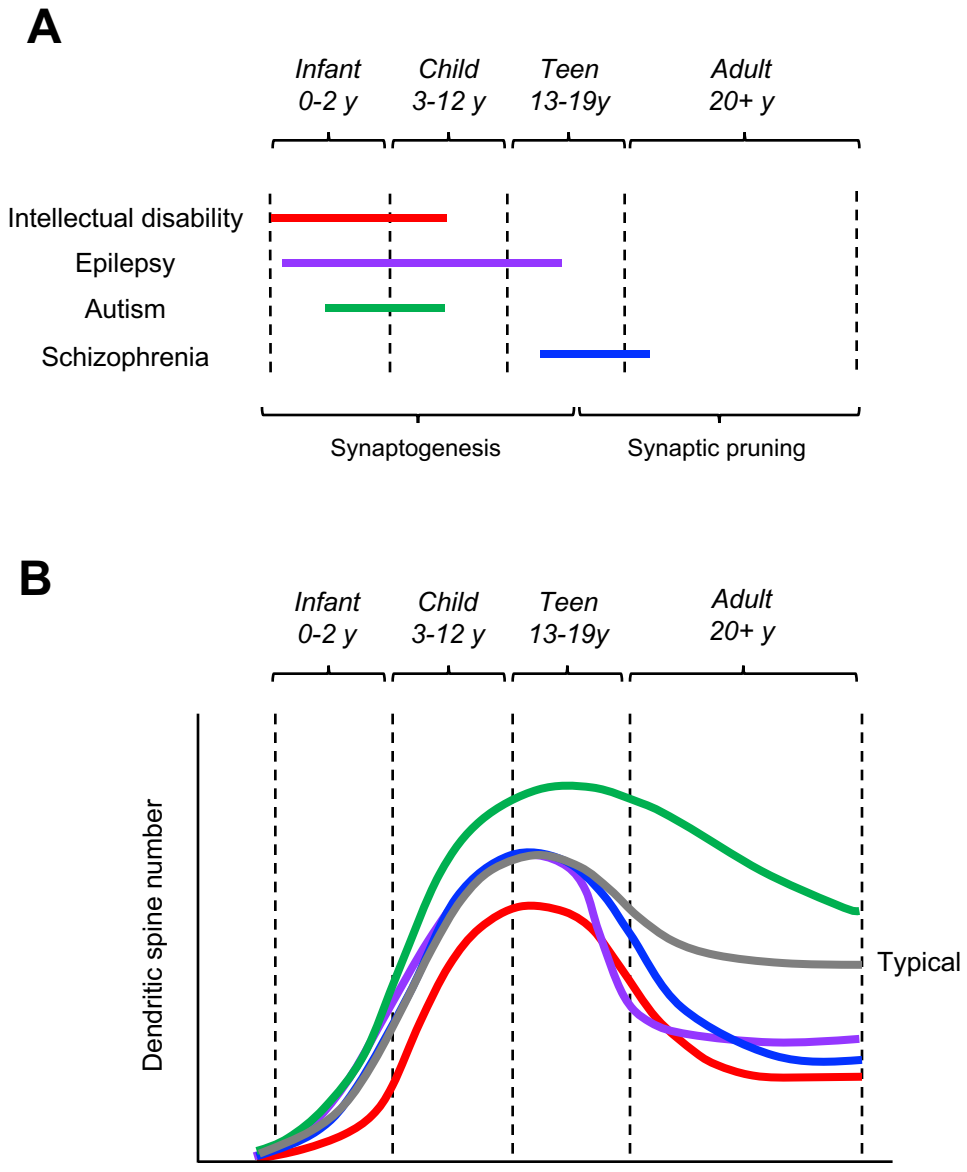


Figure 1.08. Many neurodevelopmental disorders involve synaptic aberrations

(A) Distribution in the age of onset for four common neurodevelopmental disorders. The manifestation of these diseases overlap with the milestones shown in **Figure 1.07** (e.g. synaptogenesis and synaptic pruning). This supports the idea that aberrations in specific developmental processes may underlie these illnesses. (B) In post-mortem tissue, intellectual disability, epilepsy, autism spectrum disorder, and schizophrenia have been associated with abnormal quantities of dendritic spines. Excessive spine growth or elimination during development could account for these findings. Normal development is shown in grey; disease colors are continuous between Figures A and B. Y: years. Figures adapted from Forrest et al. (2018).

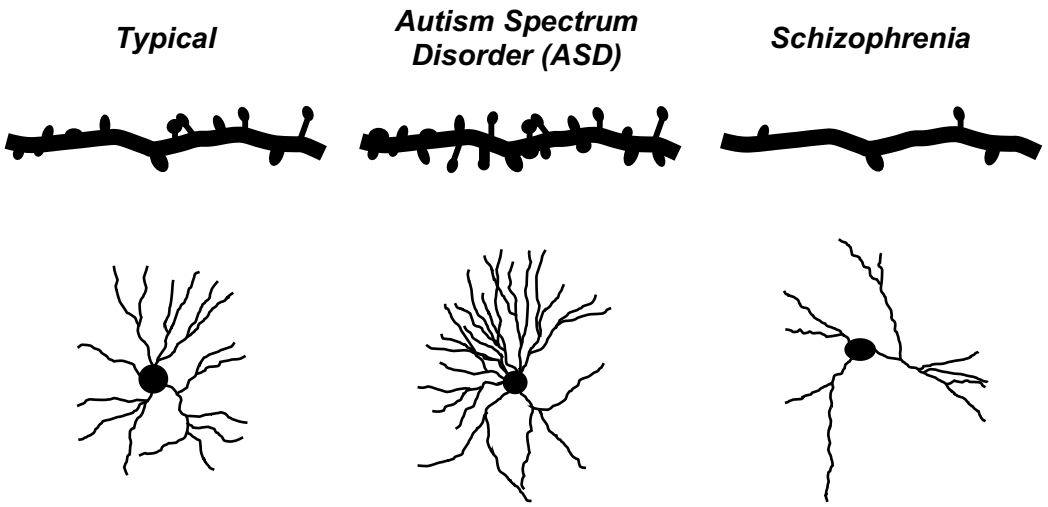


Figure 1.09. Illustration of disease-related changes to the cortical synapse

Example tracings of dendritic spines (top row) and whole neurons (bottom row) for typical development, ASD, and schizophrenia. ASD is frequently characterized by increased dendritic spine density and/or neurite branching. Conversely, schizophrenia is characterized by the opposite phenotype – decreased spine density and neurite branching. Such abnormalities may be mechanisms for disease pathogenesis.

Feature	*Autism spectrum disorder*	*Schizophrenia*
Structural imaging	Early **increases** in the volume of the frontal and temporal cortices; possible accelerated **reductions** later	**Loss** of volume in the pre-frontal and temporal cortices
Functional imaging	Altered **long distance** connectivity between the frontal and posterior lobes	Altered **short distance** connectivity in the frontal, temporal, and parietal lobes
Post-mortem synapse phenotype	**Increased*** dendritic spine density; i**mmature** appearing spines; i**ncreased** neurite branching (*Decreased spines in Rett syndrome)	**Decreased** dendritic spine density; **decreased** neurite branching; **decreased** expression of synaptic proteins

Table 1.02. Disease-related changes to the cortical synapse

Summary of changes to the neurites and dendritic spines of glutamatergic neurons in autism spectrum disorder (ASD) and schizophrenia. The two neurodevelopmental disorders typically show opposite phenotypes – one of over-connectivity in ASD, and one of under-connectivity in schizophrenia. Table adapted from Habela et al. (2015).

Gene	*Function*	*Evidence for association*	*Effect on cortical connectivity*	*Evidence for effects*
Neuroligin-3/4	Post-synaptic adhesion protein	Rare variants	Increases spine density	Cell culture
Neurexin1	Pre-synaptic adhesion protein	Rare variants	Increases spine density	Transgenic mouse
Shank3	Post-synaptic scaffold	Rare variants	Increases spine density	Cell culture
Shank2	Post-synaptic scaffold	Rare variants	Increases spine size	Cell culture

Table 1.03. Synaptic genes associated with autism spectrum disorder (ASD)

A selection of genes that regulate synaptic function in the cortex and are associated with ASD. Mutations in (most of) these genes cause increased dendritic spine density. These and other findings suggest inadequate pruning of spines may contribute to ASD. Table adapted from Penzes et al. (2011).

1.4 Unique features of the human cerebral cortex

Over the ~200 million years of mammalian evolution, the cerebral cortex has changed considerably (71). Perhaps most striking in humans is the disproportionate enlargement of the cortex relative to the rest of the brain (114). Although this increase can be observed in many mammals, it is pronounced on the human lineage (**Figure 1.10**). Regions involved in human-specialized brain functions, including language and mathematical reasoning, have expanded the most (115). These regions include the pre-frontal cortex, the parietal cortex, and the temporal cortex. The enlargement is primarily seen as an increase in cortical surface area, with only a modest increase in cortical thickness (22). Our brains are also **gyrencephalic** (convoluted), whereas those of rodents are **lissencephalic** (smooth) (**Figure 1.11**) (30). This is a consequence of the expansion of the cortex within the fixed-volume of the skull.

Four main factors are attributed to the expansion of the human cortex: 1) changes to **cell number**, 2) changes to **cell morphology**, 3) changes to **cell composition**, and 4) changes to the **duration of cortical neurogenesis**. Please refer to **Table 1.04** for an overview of key differences (adapted from Geschwind & Rakic (22)). The human cortex has approximately 1000x more neurons than the mouse cortex, and 2-3x more than the chimpanzee (**Figure 1.11**) (22). The increase in cortical thickness (approximately two-fold) is due to increases in neuron size and the amount of neuropil[2] occupying extracellular space (76). The ratio of neurons to glia has similarly increased in humans – the ratio is ~2x greater in the human pre-frontal cortex than in macaques (116). There has also been a major change in the spatial organization of neurons. Specifically, the distance between the cell bodies of neurons has become larger (i.e. neuron density has decreased), and this is particularly evident in the frontal cortex (the principal brain area for executive functions) (117). This difference

[2] Neuropil is the interwoven network of axons, dendrites, and glial cells.

– and not solely increases to neuron number – could contribute to the expansion of the human cortex. Relatedly, the expansion is seen disproportionately in upper layer neurons. Since upper layer neurons form intra-cortical connects (versus deep layer neurons that form extra-cortical connections), this could increase the integrative and computational abilities of human cortex (118). Finally, the number of cortical mini-columns, has also increased in size, number, and complexity (119, 120).

The increased size of the human cortex likely relates to changes in the neural progenitor pool. Specifically, humans have a higher proportion of proliferative progenitors compared to rodents (i.e. more that undergo self-renewing division versus those that differentiate) (77). Approximately 50-75% of the human SVZ progenitor pool are proliferative progenitors. This contrasts with what we see in rodents, where ~80-90% of the pool is neurogenic. This human-specific expansion of the SVZ progenitor pool can also be seen as an enlargement of the SVZ. In species with gyrencephalic brains – and in particular primates - the SVZ is divided into two morphologically distinct sub-zones: the **inner sub-ventricular zone (iSVZ)** and the **outer sub-ventricular zone (oSVZ)** (75, 121). The iSVZ largely resembles the SVZ of rodents, while the oSVZ is absent in most lissencephalic species. The thickness of the iSVZ remains relatively constant throughout corticogenesis, while the oSVZ grows progressively thicker as the number of progenitor cells increases.

Another key difference is the duration of cortical neurogenesis. Firstly, the onset of neurogenesis is delayed in humans compared to lissencephalic species. This implies the neuroepithelial stem cell (NSC) founder pool undergoes greater expansion before neurogenesis begins (e.g. the duration of NSC proliferation is 10x longer in humans versus rodents) (118). Neurogenesis itself is then also extended, indicating progenitors undergo greater amplification before differentiating into neurons. In humans, neurogenesis lasts ~120 days, while it takes only ~60 days in macaques and 6 days in mice (**Figure 1.12**) (22). This may, in part, explain why the human neocortex has expanded, despite the cell cycle (paradoxically) being shorter in rodents.

In that same vein, the human brain has an exceptionally extended period of development - past puberty and adolescence. Compared to all other mammals, human brains are the smallest percentage of their adult mass at birth (27). Macaques are born with brains that are 75% of their adult mass, while humans' are only 25% (27). New brain areas have also emerged on the primate lineage, in addition to the modification of pre-existing regions. For example, Broca's and Wernicke's areas, which are believed to be involved in language comprehension and production, cannot be identified in the rodent cortex. Moreover, the cortex of early mammals are believed to have had 20 cortical regions (71). In contrast, humans have over 200 distinct regions – a 10-fold increase. It is currently unknown, however, whether humans have any cell types or anatomical regions not present in other primates. An emerging view is that there is no compelling evidence to suggest they do (1).

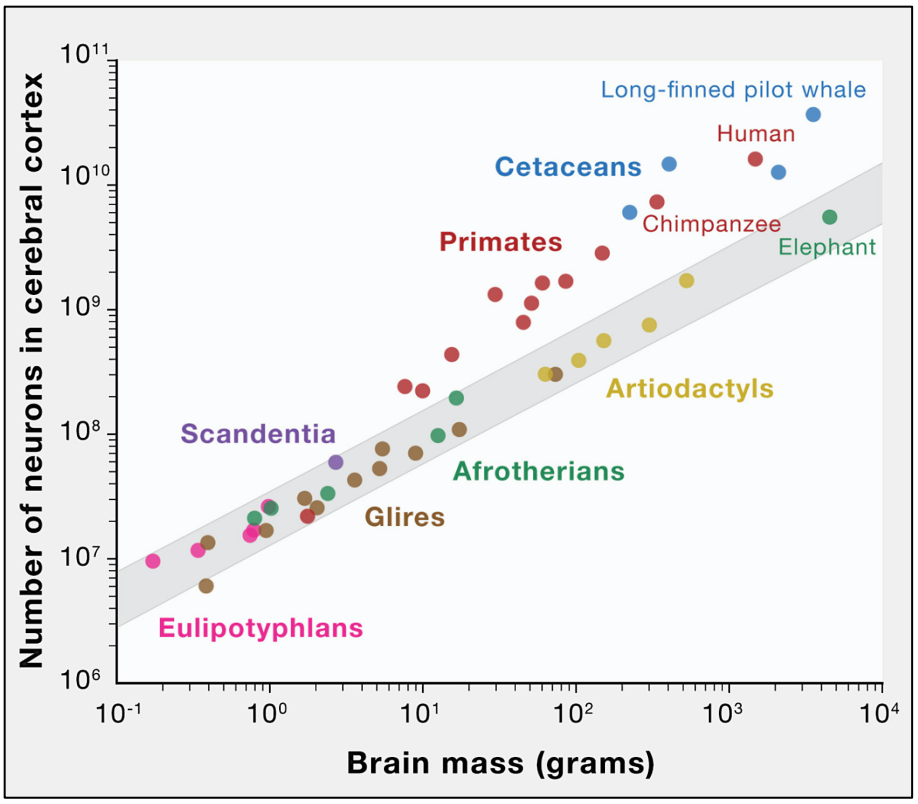


Figure 1.10. The relationship between neuronal number and brain size in mammals

In general, mammals with larger brains also have a greater number of cortical neurons – and this follows predictable scaling rules. The shaded region represents the predicted 95% upper and lower bounds (excluding cetacean and primate data). This highlights that cetaceans and primates (notably humans) have a larger number of cortical neurons than expected for their brain size. Figure taken from Sousa et al. (2017).

Cetaceans: whales, dolphins, and porpoises; **Artiodactyls:** cows, sheep, deer, and camels; **Afrotherians:** elephants, aardvarks, elephant shrews, and tenrecs; **Glires:** rodents; **Scandentia:** tree shrews; **Eulipotyphlans:** moles and hedgehogs.

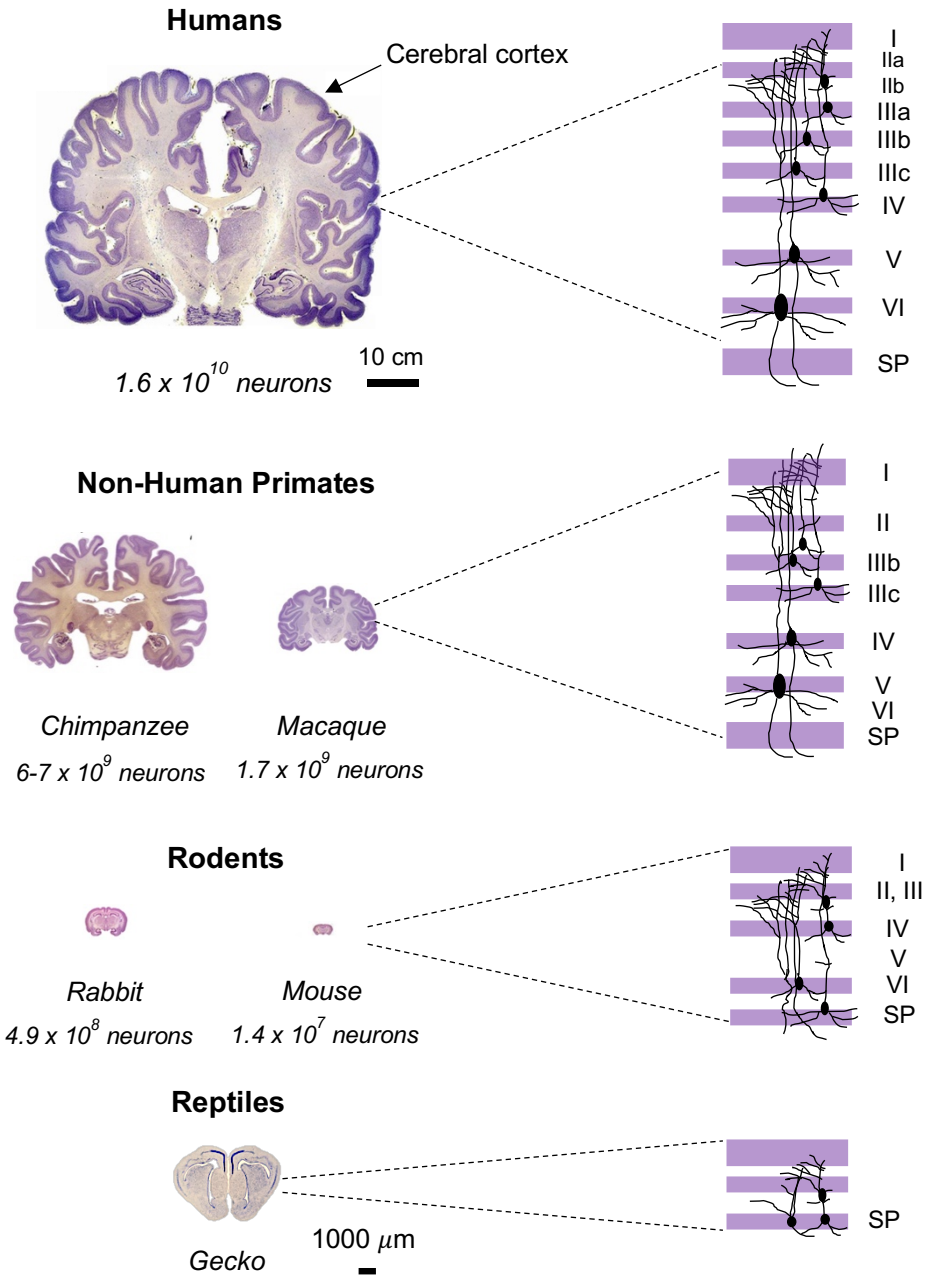


Figure 1.11. Comparison of vertebrate cortical structures

Coronal slices of human, non-human primate, rodent, and reptile cortex (*note human, non-human primates, and rodents share the same scale, while reptiles are scaled separately). The human cortex has enlarged in both thickness and surface area. Humans also have a larger number of cortical neurons, and have gyrencephalic (convoluted) brains, while rodent and reptile brains are lissencephalic (smooth). Lastly, cortical layers have progressively elaborated in mammals with larger brains. Sections taken from DeFelipe (2011). Illustration of cortical layers is adapted from Hill & Walsh (2005). Neuron counts taken from Herculano-Houzel et al. (2015). SP: sub-plate.

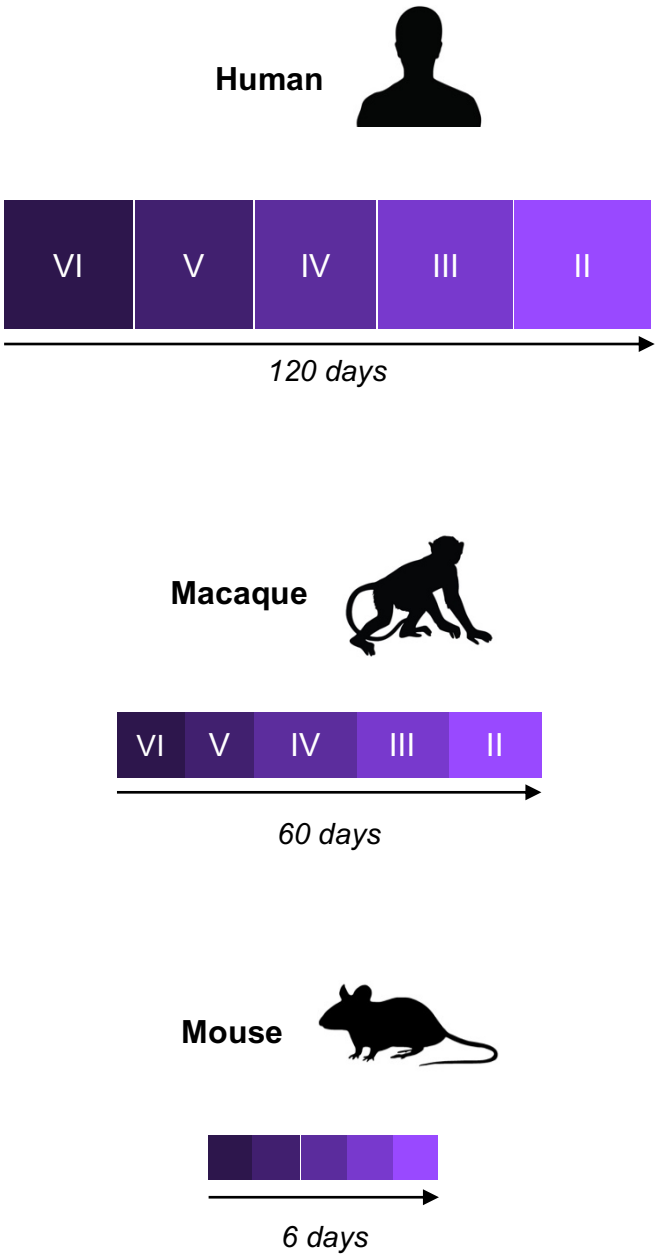


Figure 1.12. Corticogenesis is delayed and prolonged in humans

In humans, the neuroepithelial stem cell (NSC) founder pool undergoes greater expansion before the onset of neurogenesis (e.g. the duration of NSC proliferation is 10x longer in humans than in rodents). Neurogenesis itself is then also extended, indicating progenitors undergo greater amplification before differentiating into neurons. Neurogenesis lasts ~120 days in humans, while it takes only ~60 days in macaques and 6 days in mice. This prolonged development is hypothesized to underlie greater learning and plasticity in the human brain. Figure adapted from Otani et al. (2016).

Human versus mouse cortex

Quantitative Differences	*Qualitative Differences*
1000x more neurons	Gyrencephalic vs. lissencephalic
1000x increase in surface area	New genes, gene variants, expression patterns, regulatory elements
1000x more mini-columns (vertical/radial units)	New neuronal types (e.g. von Economo neurons)
Cell cycle 3-4x longer	Distinct radial glia cells
Cortical neurogenesis 20x longer in duration	Modification to cytoarchitectonic areas (e.g. A22, 28, 44, 45, 46)
Larger neuropil	Prolonged neoteny
Lower neuron density	
Higher dendritic and axonal branching	
Greater density of dendritic spines and synapses	
Enlargement of the sub-ventricular zone (SVZ)/outer subventricular zone (oSVZ)	
Increased glia to neuron ratio	

Table 1.04. Qualitative and quantitative differences between human and mouse cerebral cortex

Examples of differences in the development of human and mouse cerebral cortex. In general, these differences can be categorized as 1) changes to cell number, 2) changes to cell morphology, 3) changes to cell composition/organization, and 4) changes to the duration of cortical development. Table adapted from Geschwind and Rakic (2013).

1.4.2 Human-specific changes to synapse development, structure, and function

The finding that many 'synapse genes' are differentially expressed in human and non-human primate cortex is exciting for several reasons (26). Firstly, it falls in line with the hypothesis that increased synaptic complexity may underpin human brain function. Neurites and dendritic spines could contribute to higher cognition by increasing neuronal connectivity, integrative activity, and/or synaptic plasticity (28). Studies in the cognitive sciences have found that the coordinated firing of neurons underlies perception, emotion, and decision-making (122). More spines and/or branches could allow additional synaptic connections to be made, and information to be transferred in greater volumes, at greater speeds, or in more complex patterns.

Secondly, there are striking morphological differences between the synapses of humans and other primates. Human neurons - across several cortical regions - have longer dendrites and more elaborate dendritic branching than chimpanzees, macaques, and marmosets (28, 29, 123). Compared to chimpanzees, **total neuron length** in the human pre-frontal cortex is ~3-fold higher, **mean branch length** is ~1.4-fold higher, and **dendritic branching** is ~2-fold higher (28). The human frontal cortex also has a significantly greater number of **dendritic spines** per neuron – approximately 70% more than in chimpanzees or macaques (28, 31, 124). It is worth noting, however, that only a few comparative studies of primate neuronal morphology exist. While these studies consistently show humans have the greatest neurite length and arborization, there are conflicting accounts on whether **spine density** has increased (it almost certainly has relative to the rodent cortex, see **Figure 1.13**) (28, 32). That said, the most comprehensive study to date (124) reported humans as having the highest **number of synapses per neuron** amongst 25 primate species, including chimpanzees, bonobos, and gorillas (**Figure 1.14**).

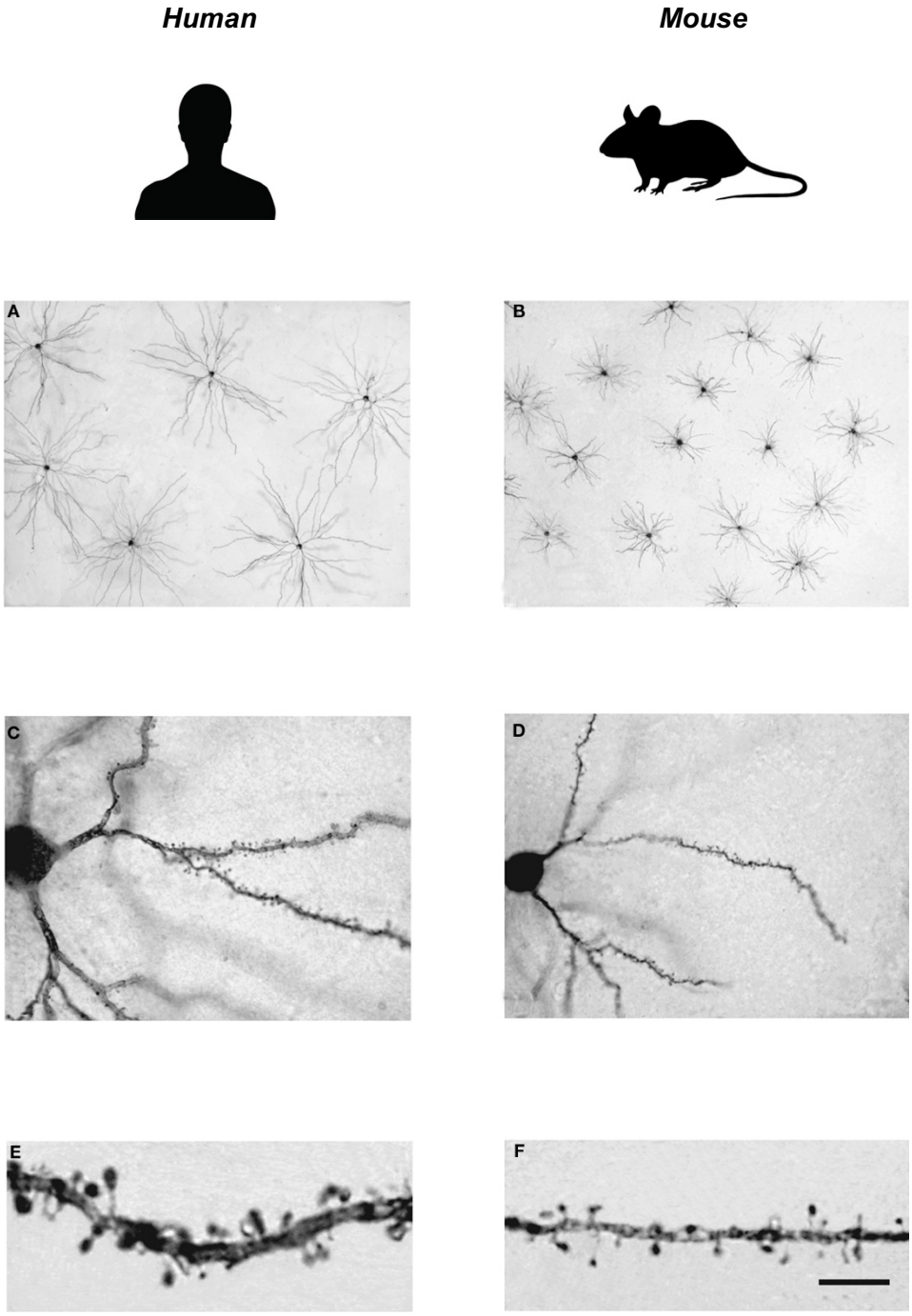


Figure 1.13. Neurites of human and mouse temporal cortex

(A-D) Human cortical neurons are larger in size than mouse neurons, and have longer neurites. (E-F) Human dendritic spines are also larger in diameter and length, and exist at higher densities than mice. Similar findings have been reported for humans in comparison to other primates. Scale bar is 425 μm in figures A-B, 45 μm in figures C-D, and 10 μm in figures E-F. Taken from Benavides-Piccione et al. (2002).

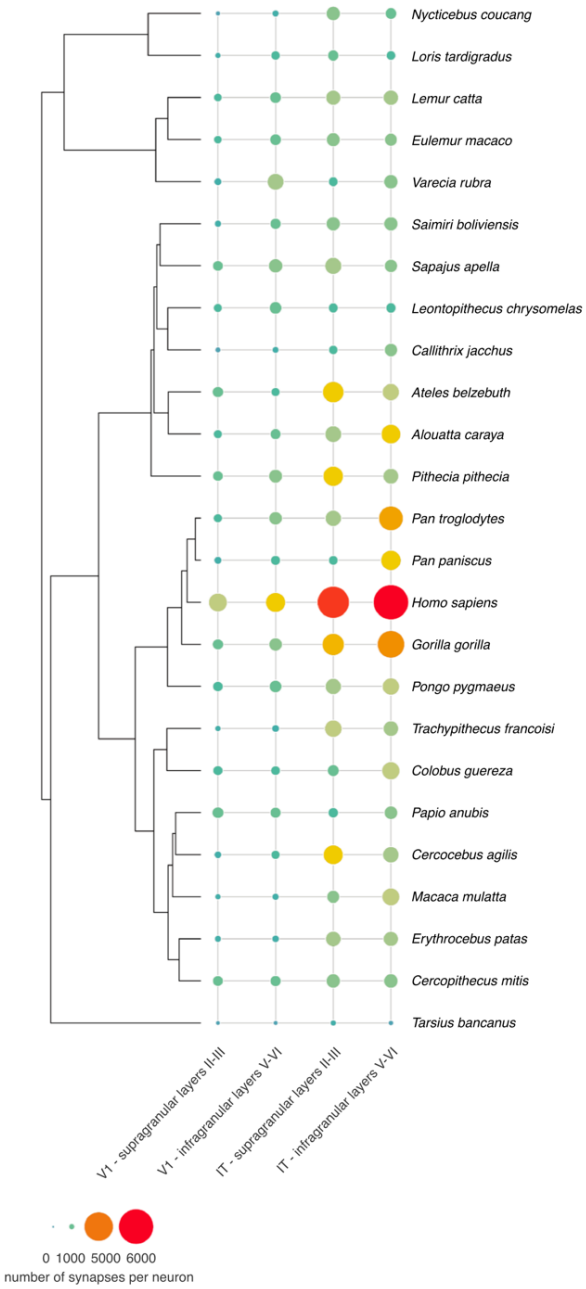


Figure 1.14. Synapses per neuron in the cortex of 25 primate species

Synapse counts are reported for the primary visual cortex (V1) and the inferior temporal cortex (IT). Samples were taken from both upper (supragranular) and deep (infragranular) cortical layers for each region. Humans have among the highest number of synapses per neuron (across the species considered). These findings support the theory of increased connectivity in the human cortex. Figure taken from Sherwood et al. (2020).

In addition to changes in neuronal structure, there have been notable human-specific alterations to cortical connectivity (1). One example is in the **arcuate fasciculus (AF)**, a tract of neurites that connects the temporo-parietal and frontal cortices (**Figure 1.15**). In humans, dysfunction of the arcuate fasciculus impairs speech production and comprehension (125). These fibers may therefore be important for language ability. AF projections in humans extend elaborately into the inferior temporal (IT) gyrus. In chimpanzees, however, projections are less extensive into the IT gyrus, and are fully absent in macaques (126). Moreover, the **superior longitudinal fasciculus (SLF)**, the main tract connecting the frontal and parietal lobes, is another set of fibres that show differences among primates (127, 128). In humans the SLF is expanded relative to chimpanzees and shows greater connectivity within the pre-frontal cortex. Fronto-parietal functions include spatial attention, social learning, and tool use – suggesting these fibers may be another important structure for human evolution.

Thirdly, there have been human-specific changes to key cell types that facilitate neurotransmission. A sub-group of excitatory neurons, called **Spindle** or **Von Economo neurons (VENs)**, are larger and more numerous in the human cortex than in other great apes (129). VENs are mainly found in the frontal lobe of big-brained species (such as primates, elephants, and cetaceans). They have long neurites, with large diameters and high conduction speeds - and may therefore be adaptations to improve information transfer in large brains (129). In that same vein, human astrocytes are bigger and more morphologically complex than those in rodents or other primates (**Figure 1.16**) (130). Crucially, astrocytes promote neurite outgrowth, as well as dendritic spine and synapse formation (131). They may also be involved in synaptic transmission and/or plasticity - although these are subject to ongoing debate (132).

Finally, synaptogenesis is delayed and extended in the human cortex (133, 134). The peak of synaptogenesis in macaque pre-frontal cortex is at 3 months of age, with synaptic pruning and axonal myelination completed by adolescence (26, 27, 134). In humans, peak synaptogenesis is reached at age 5, while pruning and

myelination continue until the third decade of life (23, 133, 134). Moreover, synaptogenesis occurs synchronously across all regions of the cortex in in macaques. By contrast, in humans it is delayed in the pre-frontal regions (135). This prolonged development is hypothesized to underlie greater learning and plasticity in the human brain (118).

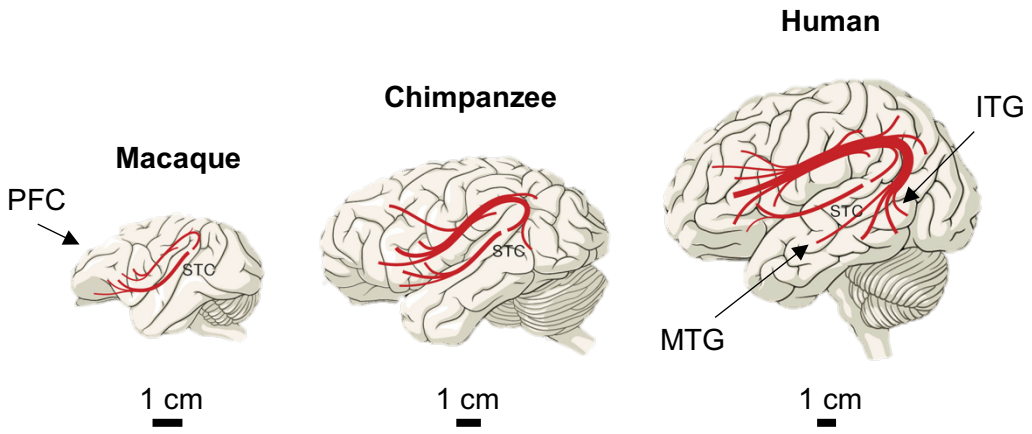


Figure 1.15. Comparison of the arcuate fasciculus (AF) track in humans and non-human primates

The arcuate fasciculus (AF) is a bundle of axons that connects the temporal and frontal lobes of the cortex. AF projections are shown in red for humans, chimpanzees, and macaques. In humans, projections extend far into the medial and inferior temporal gyri (MTG and ITG, respectively). In chimpanzees, projections to the ITG are greatly reduced, and in macaques, they are fully absent. Since human dysfunction of the arcuate fasciculus impairs speech production and comprehension, these fibers may have been important for language evolution. PFC: prefrontal cortex, STC: superior temporal cortex. Figure adapted from Sousa et al. (2017).

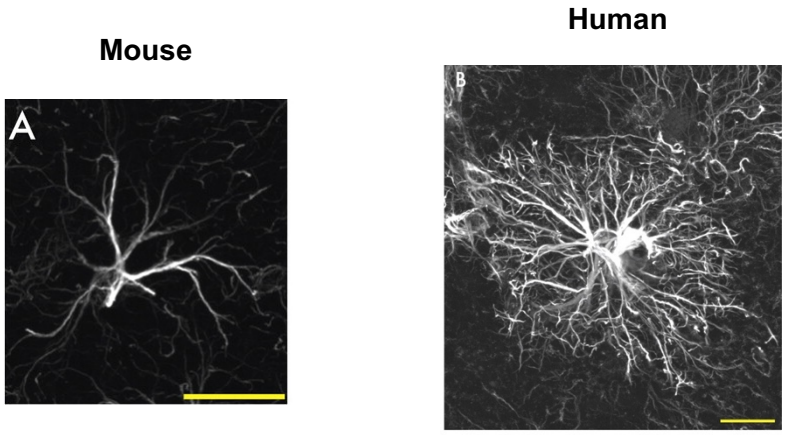


Figure 1.16. Comparison of human and mouse cortical astrocytes

Human protoplasmic astrocytes are 2.55-fold larger and have 10-fold more processes than those of rodents. Scale bar represents 20 μm. Images and data taken from Oberheim et al. (2009).

1.5 Genetic and molecular mechanisms of human cortical evolution

1.5.1 Changes to protein-coding sequences

Although changes to protein-coding sequences are unlikely to be frequent contributors to human evolution, they may still have been important factors. Perhaps the most well-known 'brain evolution gene' is the transcription factor, **forkhead box protein 2** (encoded by the ***FOXP2*** gene). Mutations in *FOXP2* lead to a dominant, Mendelian speech and language disorder (136). Individuals harbouring these mutations cannot make the fine facial movements required to produce speech.

FOXP2 is one of the most highly conserved proteins in vertebrates - it falls within the top 5% in sequence comparisons of human-rodent orthologues (137, 138). Despite this conservation, the human FOXP2 protein has two unique amino acid substitutions that are not shared by other extant primates. These substitutions were originally believed to have fixed within the last 200,000 years – around the same time that modern humans likely diverged from other hominins (23). The two amino acid substitutions are predicted to have created a novel phosphorylation site, which in turn, may have had important functional consequences for the transcription factor. In fact, subsequent studies have shown that human and chimpanzee FOXP2 have significantly different transcriptional targets (7).

Despite these initial reports, more recent research has cast doubt on the evolutionary importance of FOXP2. The first inconsistency came with the discovery that Neanderthal FOXP2 carried the same mutations as in humans (139). This suggested that the two amino acid changes must have occurred before the human-Neanderthal split, more than half a million years ago (45). In 2018, another group performed tests for positive selection at the FOXP2 locus (the Tajima's D test, which

will be described in **Chapter 7**) (140). They found that the signal that had looked like positive selection in previous studies, was an artefact caused by pooling together African genomes with those from non-African populations. Since non-African genomes were subject to a genetic bottleneck from the Out-of-Africa migrations (63), and these were pooled with only a small percentage of African genomes, a false-positive signal was detected. With more — and more high quality — genomes now available, it was concluded that there is no clear evidence of human-specific positive selection at FOXP2.

Work in animal models has largely supported that FOXP2 is important for language development. Mice possessing 'humanized' FOXP2 have accelerated learning and qualitatively different ultrasonic vocalizations from wild type (141). FoxP2 has also been implicated in vocal learning in zebra finches, a species of songbird (142). FoxP2 expression is increased in the zebra finch striatum (a part of the brain's language circuitry) during the development of language acquisition, and decreases 2 hours after singing (143, 144).

Of great importance - especially for the scope of this project - is that FOXP2 is known to both affect dendritic morphology and to regulate *CNTNAP2*. Medium spiny neurons in the striatum of humanized FoxP2 mice have significantly longer dendrites than their wild type counterparts (145). Likewise, knock down of FoxP2 in zebra finches results in reduced dendritic spine density in striatal neurons (146). A recent analysis of the transcriptional targets of FOXP2 similarly revealed they are enriched for regulators of neurite outgrowth (147). One of these targets is the candidate gene I have chosen to study, *CNTNAP2*. Thus, while the precise contribution of FOXP2 to human evolution is under review, it remains a gene important for language function. In this way, at the very least, the interaction between *CNTNAP2* and FOXP2 is still noteworthy.

1.5.2 Gene duplication and deletion

Species differences in gene copy number can arise from gene duplication, gene loss, or *de novo* gene formation. It is well known that gene duplication played a major role in the evolution of the vertebrate lineage (148). The rate of accumulation of gene duplication has increased in the African Great Apes compared to other primates, and humans are estimated to have several hundred regions of interspersed segmental duplications (149). Duplicated regions are thought to be under less adaptive constraint than their original sequences, and thus are prime material for evolution.

One example has been the multiple partial duplication events of the SLIT-ROBO Rho-GTPase Activating Protein 2 ***(SRGAP2)*** locus in the hominin lineage (150). One of these duplication events has yielded a truncated form of the protein, SRGAP2C, which is only found in humans (and no other great apes). Crucially, this protein may have increased neuronal connectivity in the human cortex. Expression of SRGAP2C in mouse excitatory neurons leads to a higher density of dendritic spines, longer spines, and longer dendritic shafts (151). Work in human and mouse neuronal cultures found that the ancestral protein, SRGAP2A, promotes the maturation of spines and slows down the migration of neurons within the developing cortex. Since SRGAP2C has opposing effects, it may therefore have played a role in the increased dendritic spine density seen in humans.

1.5.3 Alternative splicing

Alternative splicing is a powerful mechanism for generating protein diversity from a limited repertoire of genes. Two-thirds of genes in the human genome contain at least one alternative exon, and organs with many specialized cell types - such as the brain - are known to have complex alternative splicing patterns (152). Comparisons of human and chimpanzee proteins have shown approximately 6%–8% of orthologous exons have splicing level differences in corresponding tissues between the two species (153).

Functional studies of alternative splicing candidates are lacking. However, preliminary work has identified several interesting genes appropriate for further investigation. One such candidate, TAF6, is a subunit of the gene activating transcription factor, TFIID (154, 155). Isoforms of TAF6 have been implicated in apoptosis and cell cycle arrest. The alternative splicing difference between humans and chimpanzees is in the 5' untranslated region (UTR). It has been proposed that this may cause changes to human TAF6 expression, which in turn, might explain some of the differences in transcriptional profiles between humans and chimpanzees (153).

1.5.4 Epigenetic modifications

Another method to identify differences in gene expression is to compare epigenetic markers of active and inactive genes between species. These markers include histone modifications (e.g. acetylation or methylation of lysine residues) and DNA methylation of the promoter or gene body itself. For example, H3K27ac, H3K4me2, and H3K4me3 all mark active promoters and enhancers (*H3K4me3: histone H3- trimethyl-lysine 4, etc.), while H3K9me2 marks transcriptionally silent sites (156, 157). Similarly, DNA methylation can activate or repress gene transcription – the exact type of modulation depends on the location of methylation within the gene (158). Promoter methylation is associated with transcriptionally repressed genes, while gene body methylation is associated with gene activation.

Thousands of genes with human-specific activation/repression marks have been identified (38, 156-162). Many of these differences occur near genes involved in transcription regulation or signal transduction. Again, this may partially explain the differences in transcriptional profiles observed between humans and other primates. A large number of these regions are also associated with developmental and neurological functions (157). For example, the Dipeptidyl Peptidase-like 10 *(DPP10)* gene has H3K4me3 gains at its transcriptional start site in human pre-frontal cortex, but not in chimpanzees or macaques (157). *DPP10* is heavily associated with autism,

schizophrenia, and various mood disorders. It also shows evidence for selection, both in the branch separating humans from other primates and for on-going selection within human populations.

1.5.5 Non-coding RNAs

Non-coding RNAs (e.g. miRNA, siRNA, and lincRNA) provide yet another mechanism through which gene expression can be controlled. Unlike protein-coding genes, miRNAs differ greatly between humans and other species. There are nearly double the number of miRNAs in humans as there are in mice, and nearly 6x more than in drosophila (163). A noteworthy example is HAR1. HAR1 was originally identified as the highest ranked human accelerated region (HAR) by Pollard and colleagues (see **Section 1.5.6**) (9). It is a long non-coding RNA that is expressed in the cortex during weeks 7-19 of gestation - a time-point critical for neuron migration and specification (164). The human-specific mutations in the gene are believed to have critically changed the secondary structure of the non-coding RNA. HAR1 was found to be co-expressed with the gene reelin (*RELN*), which is an important mediator of cortical layering during development. HAR1 could therefore play a role in the organization of the developing cortex.

1.5.6 Human accelerated regions

As mentioned in **Section 1.1.**, the sheer size of the human genome (approximately 3 billion nucleotides) poses a major challenge for identifying the genomic regions important for human evolution. In order to prioritize sequences for further study, Pollard and colleagues (9) designed a statistical test to find **'human accelerated regions' (HARs)**. HARs are DNA sequences that fulfill two key criteria (see **Figure 1.17**) (9-11, 13, 14, 22):

i) They are **highly conserved** across a wider clade (e.g. primates, mammals, or vertebrates) - this suggests the region may be functional;
ii) They have an unexpectedly large number of **human-specific nucleotide changes** – this suggests the sequence may be important for human evolution.

For example, a 100 bp-long HAR will contain an average of ~1.7 human-specific substitutions (HARs have a mean length of 266 bp) (12). In contrast, chimpanzees (who carry the highly conserved orthologue), will have ~0.2 unique substitutions. Even if a HAR gains only a small number of human-specific changes, this rate is significantly higher than observed in other conserved elements. Moreover, the regions surrounding HARs are usually still conserved, suggesting HARs may be part of a larger functional structure (165).

Acceleration tests that are used to identify HARs compare the substitution rate **observed** on the human lineage with the rate **expected** given the rest of the phylogenetic tree (see **Table 1.05**) (12). Most studies define the 'human lineage' to include all species since our split from chimpanzees and bonobos. Only a small number of studies have separated *Homo sapiens* from archaic hominins (166). Significant acceleration of mutation rate in a locus within a lineage can indicate either 1) **positive selection** or 2) a relaxation of **negative (purifying) selection** (**Figure 1.18**). The terms 'positive' and 'negative' here refer to the directionality of the selective pressure in the population-level frequency of alleles (this will be discussed further in

Chapter 7) (167). With positive selection, beneficial alleles rise in frequency; with negative selection, harmful alleles decrease in frequency.

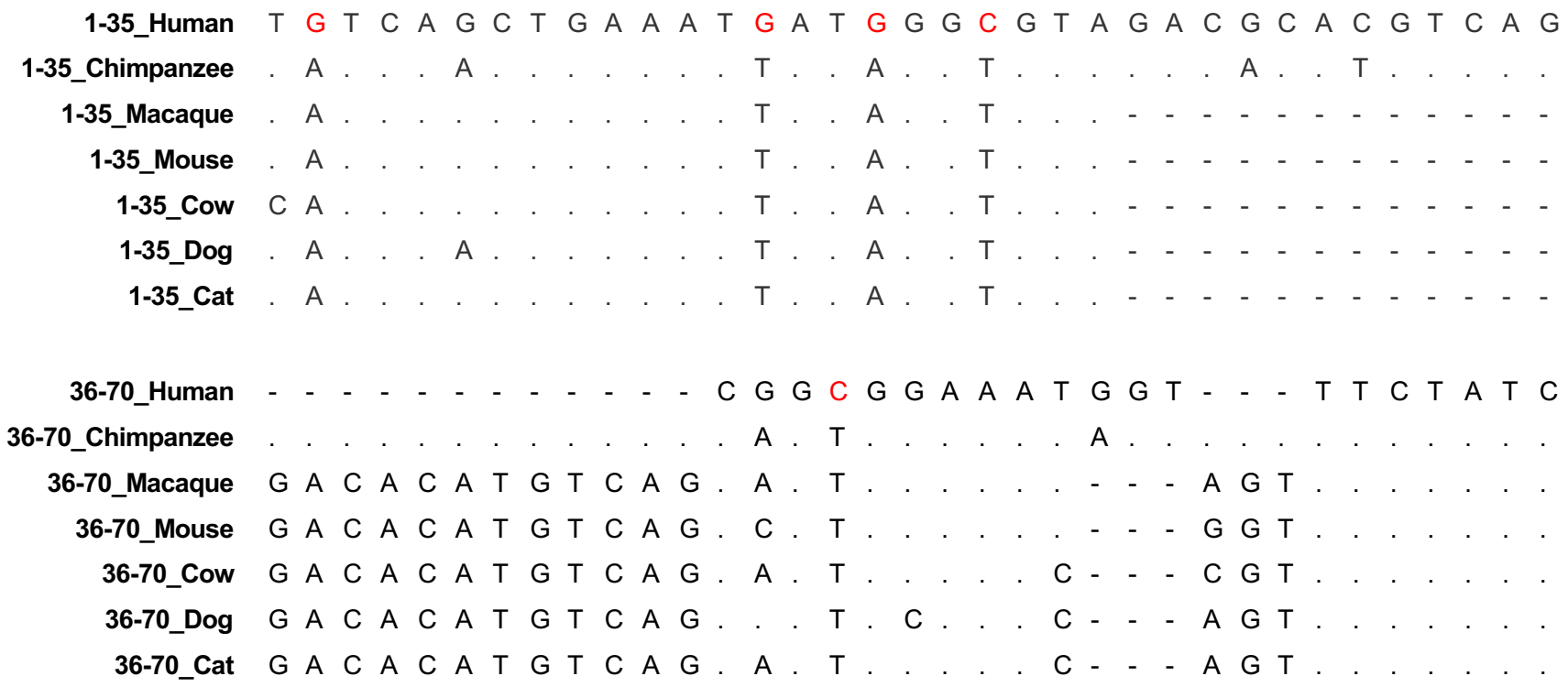


Figure 1.17. Multiple species alignment of HAR1

Human accelerated regions (HARs) are sequences of DNA that are highly conserved in other species, but have an unexpected number of human-specific substitutions. The first paper published on HARs (Pollard et al. 2006) identified HAR1 as the most accelerated region in the human genome. Only the first 70 nucleotides of HAR1 are shown for space, with 5 human-specific changes (shown in red). The full sequence is 109 bp long and contains 12 nucleotides unique to humans. HAR1 encodes a long non-coding RNA (lncRNA) expressed in the developing cortex (see Pollard et al. 2006).

HAR method	*Clade analyzed*	*Only non-coding DNA?*	*DNA alignment & conservation tools*	*Significance test*	*No. of HARs identified*	*% of HARs explained by positive selection*
HARs[a] (Pollard 2006)	17 vertebrates	No	MultiZ, PhastCons	Likelihood ratio test	202	76%
HACNSs (Prabhakar 2006)	8 vertebrates	Yes	PhastCons	Human acceleration p-value	992	Not estimated
ANCs (Bird 2007)	17 vertebrates	Yes	MultiZ, PhastCons	χ^2 relative rate test	1,356	15-19%
HARs[b] (Bush 2008)	6 mammals	Yes	MultiZ, PhastCons	Likelihood ratio test	63	Not estimated
2xHARs (Lindblad-Toh 2011)	29 mammals	No	MultiZ, PhastCons	Likelihood ratio test	563	85%
haDHSs (Gittelman 2015)	6 primates	No	Ensembl Genome Browser	Likelihood ratio test	524	70%

Table 1.05. Comparison chart of the six major human accelerated region (HAR) studies

Descriptions of each study including their methods and results. Overall, there is little overlap in the accelerated sequences discovered by each paper (see **Figure 1.21**). This is likely due to differences in methods, including the inclusion/exclusion of coding DNA, the specific species considered, as well as statistical tests, filters, and methods for species alignments. Importantly, a significant number of HARs are predicted to have arisen due to positive selection (and not GC biased gene conversion or relaxation of constraint).

HACNS: human accelerated conserved non-coding sequence, **ANC:** accelerated conserved non-coding sequence, **haDHS:** human-accelerated DNAse I hypersensitive site.

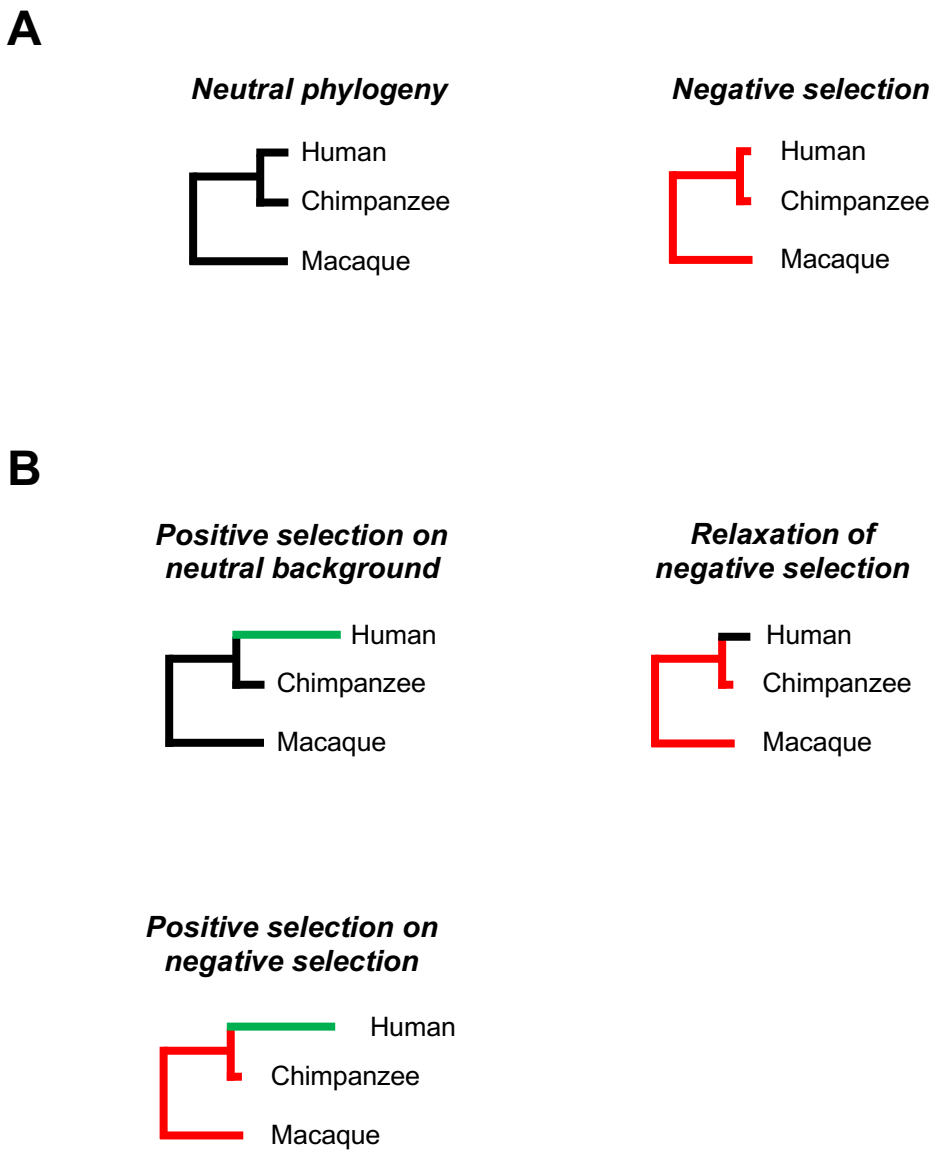


Figure 1.18. Evolutionary forces that can generate human accelerated regions

(A) Examples of phylogenies undergoing neutral evolution or negative selection. Branch lengths represent the expected number of substitutions. (B) Examples of phylogenies undergoing human-specific acceleration. The human substitution rate can accelerate due to 1) positive selection on a neutrally evolving sequence (potentially a human-specific gain of function), 2) relaxation of negative selection (potentially a human loss of function), or 3) positive selection on a previously constrained sequence (potentially a human-specific change of function). Figure adapted from Hubisz and Pollard (2014).

Acceleration tests are explicitly different from tests for positive selection. Such **'selection tests'** compare observed substitution rates to those expected under **neutral evolution.** Neutral evolution refers to changes in allele frequency due to random mutation (a process also called **genetic drift**) (168). Sequences that have neutrally evolved are neither beneficial nor harmful – they are **selectively neutral**. Since many functionally conserved elements are expected to have zero substitutions, acceleration tests can reach genome-wide significance from only a few nucleotide changes (12). As such, they can be more sensitive than those for selection (although many HARs also show evidence of positive selection) (169).

Since the discovery of HARs in 2006, approximately 3,000 HARs have been identified - 96.6% of them in non-coding DNA (170). Of these **'non-coding HARs'** (**ncHARs**), about 60% are intergenic, 35% are intronic, and the remaining are split between promoters, untranslated regions (UTRs), and genes for non-coding RNAs (ncRNAs) (12). This distribution has made it difficult to assign HARs a function, because most of the non-coding genome is uncharacterized. That said, a landmark study from Capra et al. (170) has provided strong evidence that many HARs are **gene enhancers**. The authors used existing functional genomics data, in combination with machine learning algorithms, to show that 60% of ncHARs overlap epigenetic enhancer marks like H3K4me1, H3K27ac, or p300. Half of these were predicted to target genes active during development, and one third were thought to act in the brain (**Figure 1.19**)(170). These predictions are supported by the finding that HARs are highly enriched in transcription factor binding motifs (2, 170). ~60% of HARs are also located within 1 MB of a gene that is differentially expressed between humans and chimpanzees (171). While only a few HARs have been experimentally characterized (most using mouse enhancer assays), these studies have generally validated the predictions of Capra and colleagues (171). Such findings also support the theory that changes to gene expression (strength, location, or timing) have been key to human evolution (**Figure 1.20**).

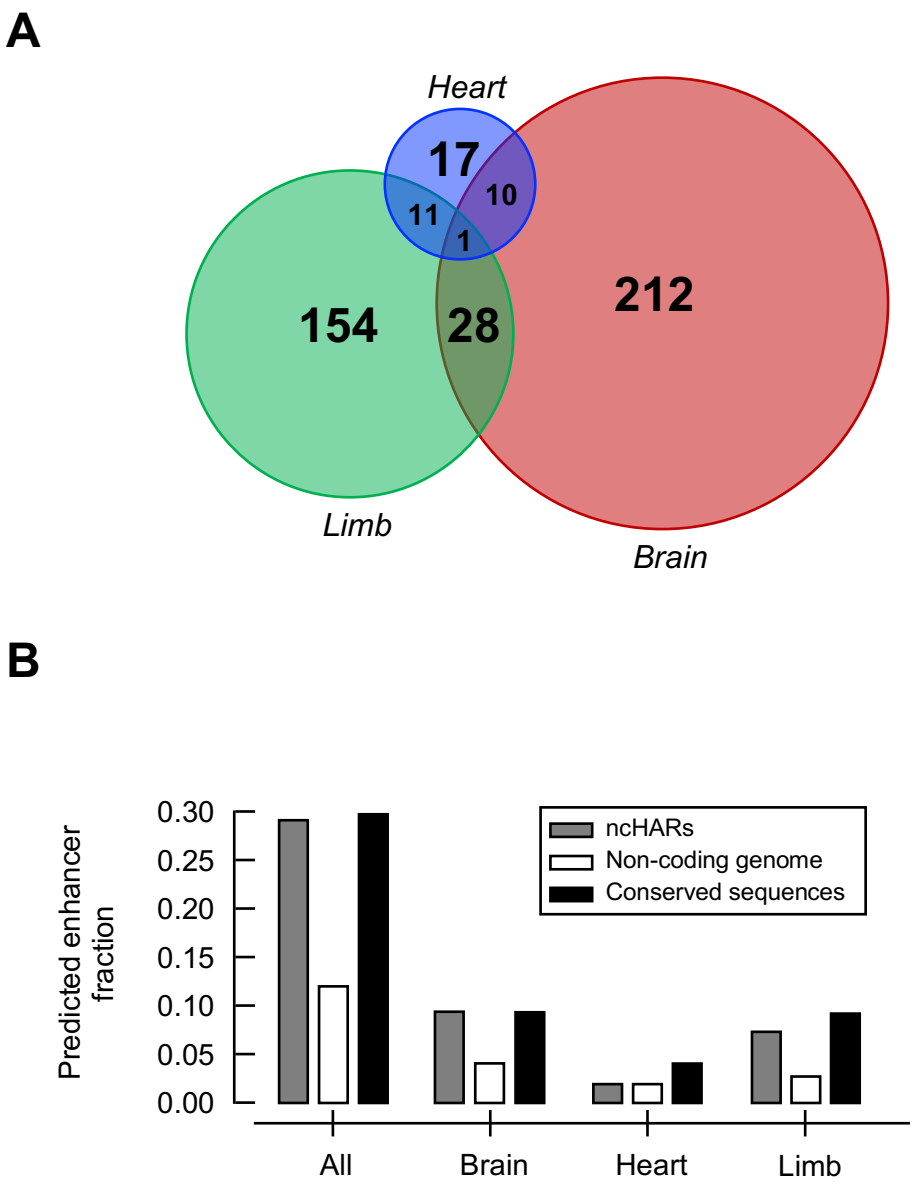


Figure 1.19. Predicted ncHAR enhancers and their tissues of activity

(A) Capra et al. (2013) applied an enhancer prediction pipeline to the 2649 HARs located in non-coding DNA (referred to as 'ncHARs'). 773 were predicted to be developmental enhancers, including 251 brain enhancers, 194 limb enhancers, and 39 heart enhancers. (B) The authors also compared the ncHARs predicted to be enhancers with the genome-wide non-coding background, and with non-coding conserved regions (mammalian phastCons elements). The ncHARs are significantly enriched for predicted enhancer activity compared with the genomic background (particularly in putative brain and limb enhancers). However, they are not significantly different from conserved non-coding regions.

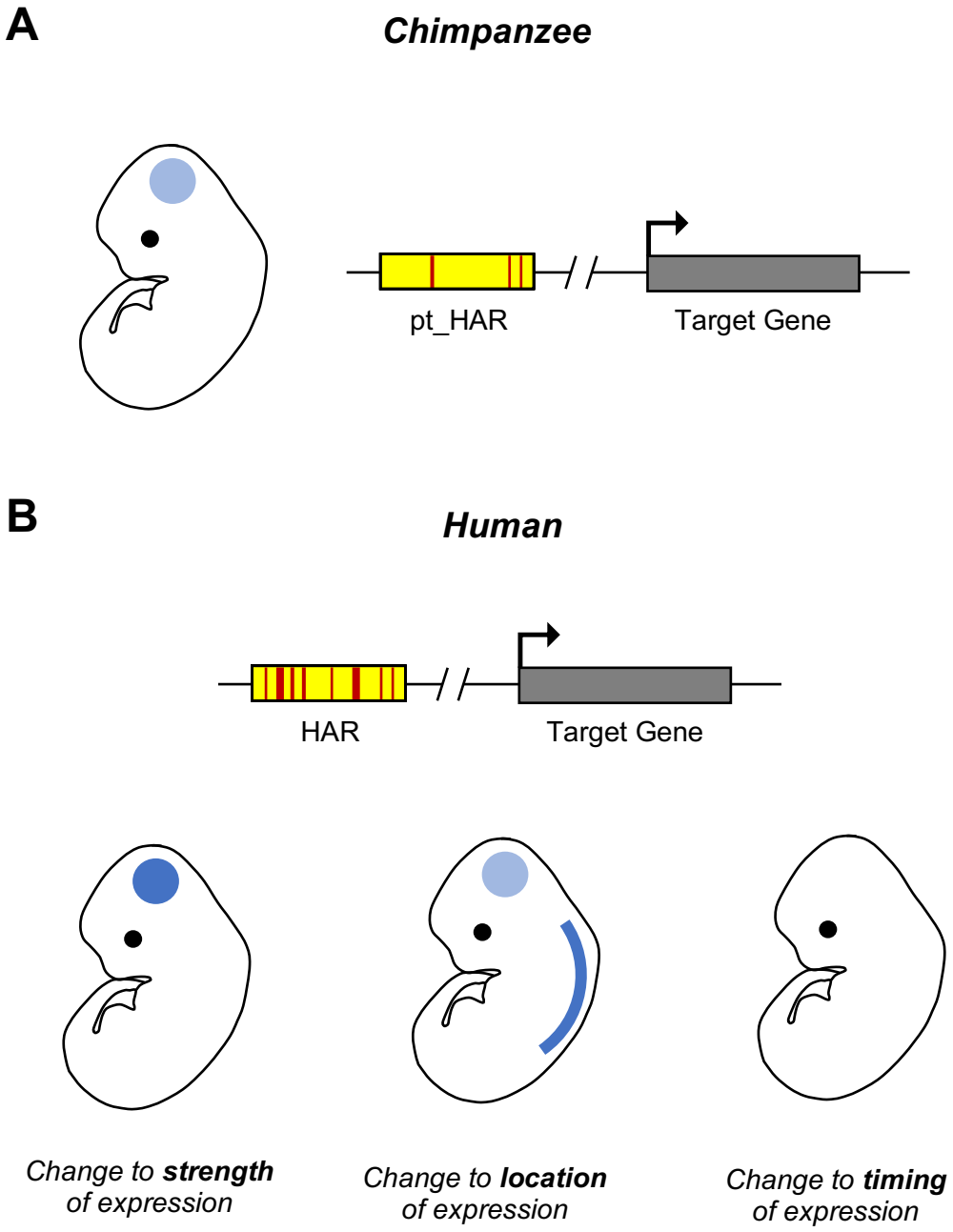


Figure 1.20. ncHARs may be human-specific gene enhancers

Example of a target gene under the control of a HAR-containing enhancer. Red lines in the HAR/enhancer sequence (yellow box) represent species-specific nucleotide substitutions. In the chimpanzee embryo (A) the target gene is weakly expressed in the developing cortex. In the human embryo (B) the presence of the HAR has altered the activity of the enhancer to change either the strength, location, or timing of the target gene's expression. Figure adapted from Boyd et al. (2015).

One notable HAR is **HARE5** (human accelerated regulatory enhancer 5), which has been identified as an enhancer of the Frizzled 8 gene *(FZD8)* (172). *FZD8* is a receptor for the *Wnt* signaling pathway that is crucial for cortical development (173). Chromosome conformation capture assays revealed that HARE5 physically contacts the *Fzd8* promoter in the embryonic mouse cortex. Boyd et al (172) subsequently generated transgenic mice with *Fzd8* under the control of either the human HARE5 sequence (HARE5) or the chimpanzee equivalent (pt_HARE5). HARE5 was found to drive early and strong expression of *FZD8* at the start of corticogenesis, coinciding with when *FZD8* is expressed in neural progenitor cells (172). More significantly, mice with HARE5 had an accelerated neural progenitor cell cycle and consequently increased brain size. This result was not observed in mice expressing pt_HARE5 (who were indistinguishable from wild type).

Relatedly, HARs are important for their potential involvement in neurodevelopmental disorders. Not only are HARs commonly located in genes linked with ASD and schizophrenia (2, 174, 175), but many of these genes are thought to be dosage sensitive. This suggests a mechanism whereby mutations in HARs lead to a loss of enhancer function, and therefore to pathological gene expression and ultimately disease. One study estimated that point mutations in HARs could account for up to 5% of consanguineous ASD cases (2). Moreover, in autistic individuals with *de novo* copy number variants (CNVs), a significant proportion involve CNVs that overlap HARs. Lastly, HARs are enriched in gene ontology (GO) terms associated with potential disease-causing cellular processes, including synapse formation, GABA/glutamate metabolism, and neurite outgrowth (175).

HARs and their role as developmental enhancers may be important contributors to human evolution and disease. However, there are a few important caveats worth mentioning. Firstly, while most HAR substitutions are fixed in humans, approximately 10% were subsequently found to be polymorphic (12). A sequence that is important for human evolution would be expected to be shared by all humans. Secondly, across the six major HAR studies, there is poor overlap between the

sequences identified as accelerated (**Figure 1.21**). Although this discrepancy is likely explained by their use of different methods, the lack of overlap is still disappointingly low. Thirdly, others have argued that HARs are faulty approximations for positive selection. Acceleration tests may be picking up the relaxation of negative selection (loss of constraint), which is not adaptive evolution. They may also be detecting increased substitution caused by **GC-biased gene conversion (gBGC)** (169). gBGC occurs through DNA mismatch repair during meiotic recombination. It results from a bias towards guanine or cytosine bases (which form strong bonds), away from adenine or thymine bases (which form weak bonds). This leads to a high proportion of A/T to G/C substitutions, and a high G/C content in affected regions. Therefore, gBGC can lead to false positives on tests of acceleration.

Subsequent analyses have confirmed that the majority of HARs (~76%) are best explained without the effect of gBGC. About half of HARs (~55%) have substitution rates that significantly exceed the rate predicted under neutrality, arguing they underwent positive selection. ~21% are consistent with a relaxation of constraint, and a substantial minority (~19%) have likely been generated from gBGC alone (169). HARs created by loss of constraint or by gBGC, however, may still have led to human-specific phenotypes through loss-of-function. For example, HAR2 is thought to have lost its repressor function (that is still present in other species) by a gBGC event, which led to a gain of enhancer activity in humans (176, 177).

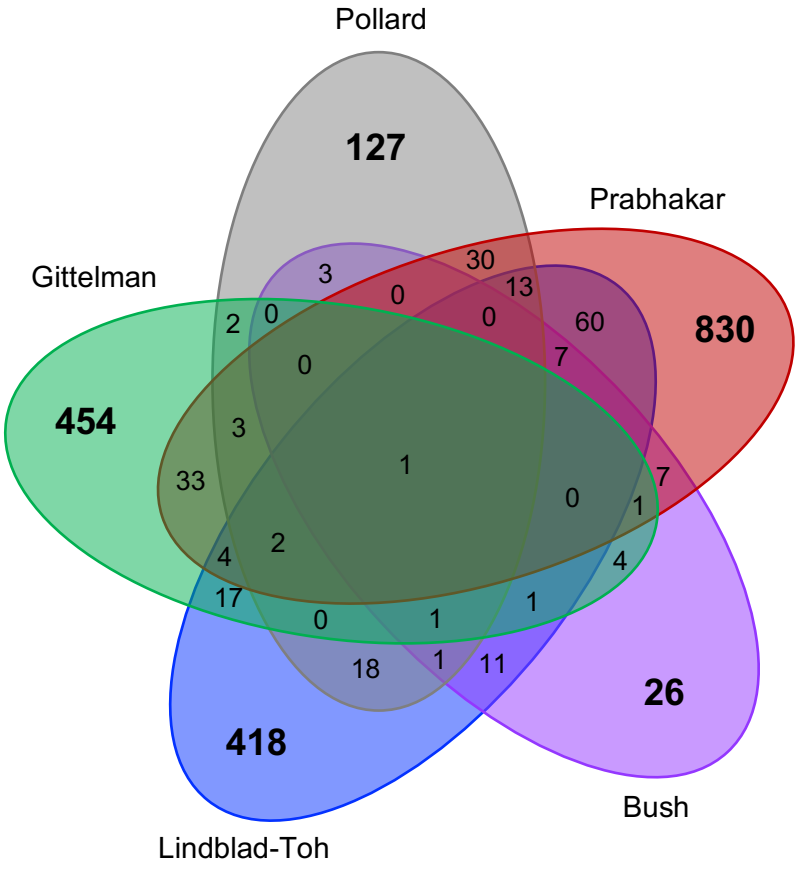


Figure 1.21. Overlap of ncHARs from five major HAR studies

The ncHARs identified by five major studies show only modest overlap (Bird et al. is not included). This is likely due to differences in the methods used to identify accelerated sequences. Prabhakar and Bird have the greatest number of overlap between the studies (241 shared HARs). Only one HAR is shared by all five studies. Figure adapted from Gittelman et al. (2015)

1.6 PSC-derived models of the human cerebral cortex

1.6.1 Current models of human brain development/disease

As mentioned, access to primate cortical tissue is heavily limited for ethical reasons. In humans, brain tissue is primarily taken from deceased donors, meaning living processes cannot be studied and sample sizes are generally low. There is also the potential for confounds based on both pre- and post-mortem conditions. For example, changes to the brain's pH, blood oxygenation, and perfusion at the time of death can all influence tissue quality and its ability to provide accurate results (178). Likewise, the **post-mortem interval (PMI)**, the time between death and tissue freezing/fixing, can lead to further biochemical changes in the brain (179). Other drawbacks of primary tissue use include that functional studies cannot be performed, and diseases can typically only be studied at advanced stages (potentially blurring cause and consequence). Post-mortem samples, however, are the main opportunity to study primary human brain tissue. As such, they remain an invaluable resource.

Animal models, primarily rodents, serve as another important tool for studying human brain development/disease. These models are particularly useful as 'whole organisms', which create more realistic cellular environments and allow testing of behavioural phenotypes. Functional studies are possible, although they may be long, laborious, and costly (5). Perhaps their greatest drawback is that they are non-human, having split from our lineage approximately 75 million years ago (180). This divergence means they may not accurately model human processes – particularly so for human-specific phenotypes.

1.6.2 The Shi et al. (2012) protocol for human PSC-derived cortical neurons

In 2012 our lab generated a protocol (herein referred to as the **'Shi et al. protocol'**) that produces human forebrain neurons from pluripotent stem cells (PSCs) (173, 181). This strategy uses **retinoic acid** and **dual SMAD inhibition** to direct PSCs to an anterior neuroectodermal fate (see **Figure 1.22**, and **Chapter 3 Methods** for full details) (181). Retinoic acid (a derivative of vitamin A) has been shown to be critical for cortical induction *in vivo* and *in vitro* (181, 182). Additionally, two small molecules, **dorsomorphin** and **SB431542**, are used to inhibit the SMAD signalling pathway (181). SMAD operates downstream of the **TGFβ/nodal** and **BMP** proteins that 1) maintain pluripotency and 2) promote mesodermal, trophoectodermal, and endodermal fates during PSC differentiation (183, 184). Blocking SMAD signalling, therefore, induces neuroectoderm formation by removing pluripotency and by suppressing alternative cell fates. The specified neuroectoderm will then spontaneously develop into telencephalon (without the need for any additional exogenous factors) (181, 185).

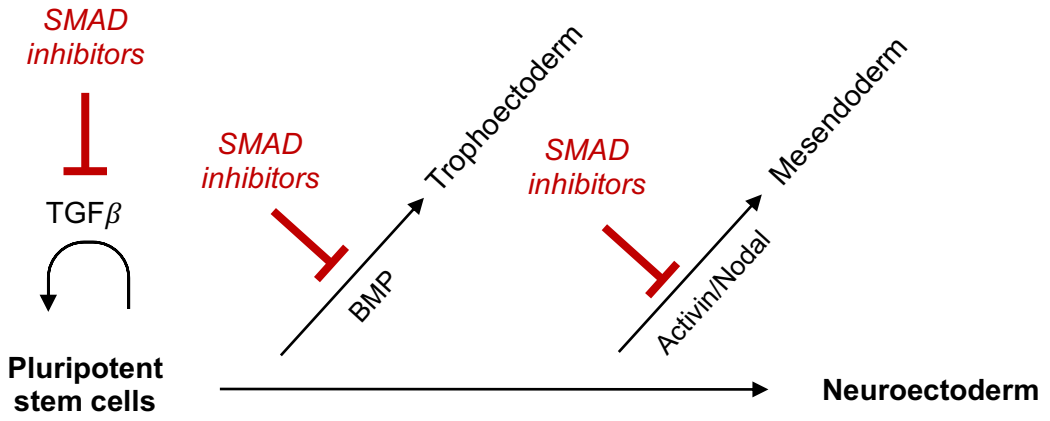


Figure 1.22. Dual SMAD inhibition

Two small molecules, dorsomorphin and SB431542, are used to inhibit SMAD signalling. SMAD operates downstream of the TGFβ/nodal and BMP proteins that 1) maintain pluripotency and 2) promote trophoectodermal, mesodermal, and endodermal fates during PSC differentiation. Blocking SMAD signalling, therefore, induces neuroectoderm formation by removing pluripotency and by suppressing alternative cell fates. Figure adapted from Chambers et al. (2009)

The entire induction process, where retinoic acid, dorsomorphin, and SB431542 are added to PSCs via culture media, takes only 12 days to complete (see **Figure 1.23**). The resulting neural progenitor cells are then treated with **FGF2** for four days to promote progenitor self-renewal, and ultimately, a larger neuronal output (181). All types of neural progenitor cells are formed, including neuroepithelial stem cells, radial glia, and intermediate progenitors (173). Deep layer neurons appear after 2-3 weeks, and layer II neurons (the last cortical layer to form) around day 90 (173). These PSC-derived neurons are functionally active, with synapses present from day ~45 onwards, and full neural networks between 2-3 months (173, 186). Following the completion of neurogenesis (around day 100), neural progenitors then commence gliogenesis to produce astrocytes. The vast majority of the cells produced by this protocol are glutamatergic cortical neurons. Based on single cell RNA-Seq conducted by myself and others in the lab (see **Chapter 5** for full details), it is also known that small populations of cortical hem, choroid plexus, and inhibitory cortical neurons are also produced. While there is some variability in induction outcome (see Strano et al. (187) for a full description), all inductions are quality-checked using a transcriptomic analysis outlined in **Chapter 4.2**.

The Shi et al. protocol faithfully recapitulates human corticogenesis as observed *in vivo*. Cells are formed in the same temporal order (i.e. deep layer neurons before upper layer), and over a comparable timeframe (181). They also display similar morphology, gene expression profiles, and cellular activity to their primary counterparts (173, 181, 186). Critically, species differences in cortical development appear to be similarly preserved. For example, it is possible to apply the Shi et al. protocol to PSCs from non-human primates (see Otani et al. (4)). These experiments revealed that neurogenesis in our human *in vitro* cortical cultures lasts approximately twice as long as in PSC-derived cultures from macaques - matching the expectations *in vivo* (4, 188). Human progenitors also undergo more rounds of self-proliferation than macaques, and develop synaptic activity later in development (4).

PSC-derived forebrain neurons are an extremely valuable model to investigate hypotheses about human brain evolution. Firstly, the use of human cells avoids inter-species differences - it is a human system with which to explore uniquely human aspects of the brain. Since changes to cortical development are believed to underpin human cognitive evolution, the ability to observe 'corticogenesis in a dish' is crucial to identifying important changes to the human developmental programme. For obvious ethical reasons, it is not possible to study human brain development *in vivo.* A similar rationale applies to studying disease pathogenesis. The use of PSC-derived cells allows the very earliest stages of disease to be investigated, and for disease progression to be examined over time. Finally, PSC-derived cultures can be functionally experimented on in large scales and with relative ease (5).

As with all model systems, there are caveats to the use of PSC-derived cortical cultures. The Shi et al. protocol efficiently generates excitatory cortical neurons, however, other neuronal and non-neuronal cell types that normally accompany *in vivo* corticogenesis are not produced. For example, cortical inhibitory neurons and oligodendrocytes are typically absent from our cultures (181). Additionally, 2D cortical cultures do not adopt the typical structural organization of the embryonic cortex. There are no defined VZ, SVZ, or cortical layers - instead, cells of different sub-types coexist at random. While 3D **'cortical organoids'** are a potential workaround, they are not currently practical for all types of studies. This is particularly the case for investigations of neuron morphology, since their 3D structure makes them difficult to image. For these and other reasons, my PhD project has relied on the Shi et al. protocol to produce 2D forebrain cultures. Wherever possible, both human and macaque cell lines were used to study the development and evolution of cortical connectivity.

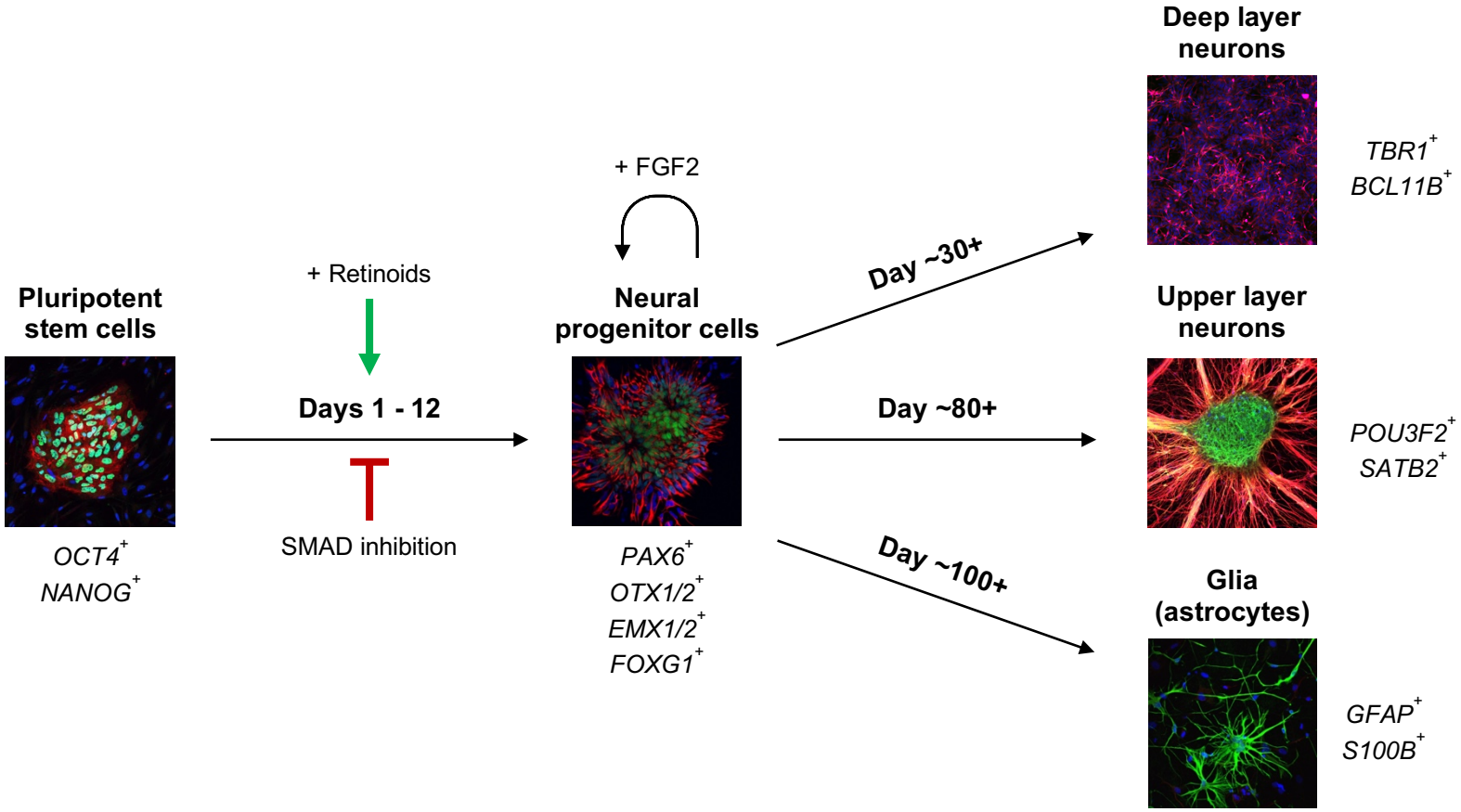


Figure 1.23. Overview of the Shi et al. (2012) protocol

Human pluripotent stem cells are directed to a forebrain fate by a combinatorial retinoid activation/dual SMAD inhibition strategy. All timeframes are shown relative to the start of the induction process (day 0). Neural induction occurs during the first 12 days, which is marked by a transition from pluripotent stem cells to neural progenitor cells. Exogenous FGF2 is added between days 13-16 to promote progenitor self-renewal and prevent neural differentiation. After this, cells are kept in neural maintenance media and allow to spontaneously develop into forebrain neurons. Deep layer cortical neurons are produced around days 20-30, followed by upper layer neurons at day ~80. After the bulk of neurogenesis has completed, neural progenitors begin producing astrocytes. Genes expressed by each induction stage are shown in italics. Cell images were kindly provided by Dr. Steve Moore of the Livesey lab.

1.7. Thesis aims

With all this in mind, I focused my PhD on the interface between neurodevelopmental disease research and brain evolution studies. Specifically, I conducted experiments to begin to dissect out: i) the precise role of 'Contactin-associated Protein-like 2' *(CNTNAP2)* in human brain evolution; and ii) how mutations in *CNTNAP2* cause human-specific diseases like ASD, SLI, and ID. Using a combination of computational and experimental approaches, I aimed to answer the following two key questions:

1) *How do changes to **CNTNAP2 expression** in human cortical neurons affect the development of **neurites** and **dendritic spines**?*
    a. *Does this affect synaptic and network function?*

2) *What are the **functions** of the **HARs** within CNTNAP2?*
    a. *Are they **enhancers** (of CNTNAP2)?*

Answers to these questions can potentially shed light on what makes us human. Critically, they may also inform on the causes and potential treatments of diseases associated with *CNTNAP2* (22). These illnesses affect millions of individuals across the globe, and cause life-long dependency (189). A detailed introduction to the *CNTNAP2* gene is provided in **Chapter 2**. This is followed by materials and methods in **Chapter 3**, and five subsequent chapters:

- **Chapter 4:** an investigation of *CNTNAP2* expression in human and macaque *in vitro* forebrain neurons,
- **Chapter 5:** single cell RNA sequencing studies of *CNTNAP2* WT and KO human cortical cultures,
- **Chapter 6:** modelling *CNTNAP2* loss-of-function in human stem cell-derived forebrain systems

- **Chapter 7:** evolutionary and functional studies of the *CNTNAP2* HARs
- **Chapter 8:** conclusions and future directions.

Chapter 2

The *CNTNAP2* Gene

2.1 The *CNTNAP2* gene and CASPR2 protein

The **contactin-associated protein-like 2** gene ***(CNTNAP2)*** is located at chromosome 7q35 (see **Figure 2.01**). *CNTNAP2* spans 2.3Mb across 24 exons, and is the one of the largest genes in the genome (44). Mutations in the gene have been reported to cause human-specific neurological disorders like autism spectrum disorder (ASD) and specific language impairment (SLI) (see **Section 2.3**). There are 25 transcripts of *CNTNAP2*, only four of which are protein coding (Ensembl release 99 – January 2020) (see **Figure 2.02**). Transcripts are designated as protein-coding by Ensembl if they contain an open reading frame. Coding transcripts 201 and 207 have been detected in chimpanzee pre-frontal cortex, transcript 201 in gorilla, and both 201 and 202 in macaque (38). Transcript 201 is the canonical and most abundant isoform.

CNTNAP2 encodes the **contactin-associated protein-like 2 (CASPR2)** protein. CASPR2 is a single-pass transmembrane protein composed of 1331 residues and with a mass of 138 kDa (43). CASPR2 belongs to the neurexin superfamily of transmembrane proteins, which are cell adhesion molecules that are involved in synapse formation and function. The CASPR2 protein contains eight extracellular, one transmembrane, and two intracellular domains. The extracellular domains include a discoidin domain and a fibrinogen-like domain, both of which are known to facilitate cell-cell adhesion and interactions with the extracellular matrix (44). The remaining extracellular domains are four laminin-G domains and two epidermal growth factor-like domains. These are predicted to mediate receptor-ligand interactions and cell adhesion, migration, and differentiation. The intracellular region of CASPR2 is mostly involved in protein-protein interactions, as it contains a type II PDZ domain and a protein 4.1B binding site.

The CASPR2 protein is highly conserved amongst mammals – for example, human and mouse amino acid sequences are 94% identical (43). This conservation is even greater between humans and chimpanzees with only 6/1331 residues differing (99.5% identity). Comparisons with archaic humans show yet further conservation. Between both Neanderthals and Denisovans, only one or two amino acid differences are noted with modern humans, respectively (43). Residue 345 is a valine in *Homo sapiens* but isoleucine in the two other species (this residue is located within a laminin-G domain). However, it is unclear whether this amino acid change has had functional consequences. Rodents also contain a valine in this position, and isoleucine shares many similar biochemical properties to valine. One other noteworthy difference exists: position 215 is an asparagine in humans and Denisovans, but in all other species (including Neanderthals) it is a Lysine. PolyPhen-2 mutation prediction software calculate both 345-I and 215-N as functionally neutral (190).

Although the molecular function(s) of CASPR2 are not completely understood, it was first characterized in the axon initial segment (AIS) and juxtaparanodal regions of myelinated neurons (191). At juxtaparanodes, CASPR2 forms a complex with the contactin-2 protein (CNTN2). This complex is required for the clustering of voltage-gated K^+ channels which function in the conduction of nerve impulses (192). Several studies have reported *Cntnap2* knockout mice have a significant reduction in the density of Kv1.2 potassium channels in cortical myelinated axons (193, 194). Scott and colleagues (194) observed these changes correlated with an increase in excitatory transmission and increased probability of neurotransmitter release. However, given that *CNTNAP2* expression is high in development - at timepoints prior to nerve myelination - other developmental functions are also likely to exist. These include proposed roles in neurite outgrowth and the formation of synaptic connections (discussed in **Section 2.5**).

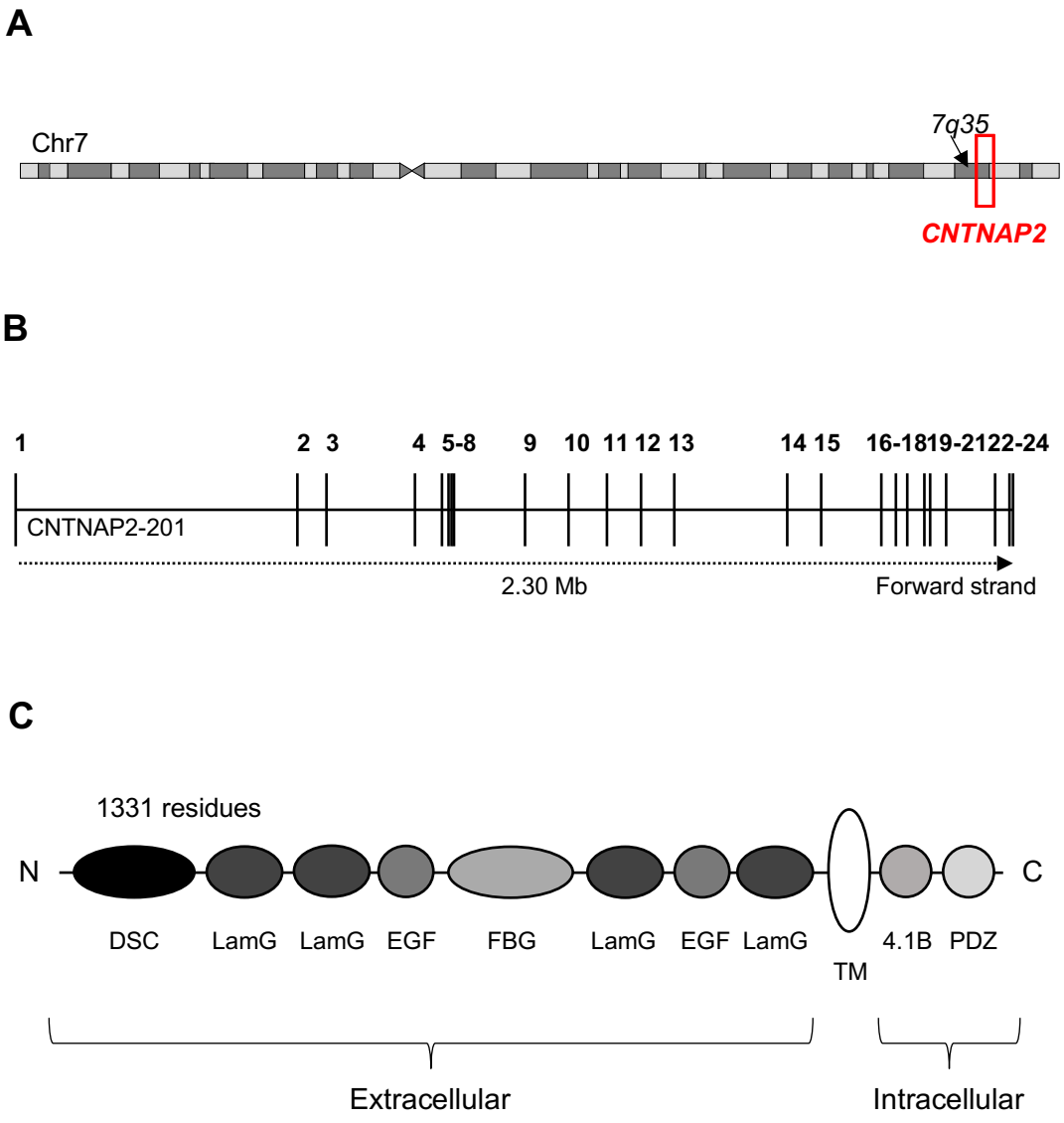


Figure 2.01. Location and structure of the *CNTNAP2* gene

(A) *CNTNAP2* is located at the distal end of the long arm (q) of chromosome 7 (GRCh38 chr7: 146,116,002 - 148,420,998). The locus occupies approximately 1.6% of chromosome 7. (B) Schematic of the 24 exons (numbered) and 23 introns of the *CNTNAP2* gene. Transcript *CNTNAP2-201* is the canonical transcript. (C) Structure of the CASPR2 protein and its 11 domains. N-terminus (N); discoidin domain (DSC); laminin-G domain (LamG); epidermal growth factor-like domain (EGF); fibrinogen-like domain (FBG); transmembrane domain (TM); type II PDZ domain (PDZ).

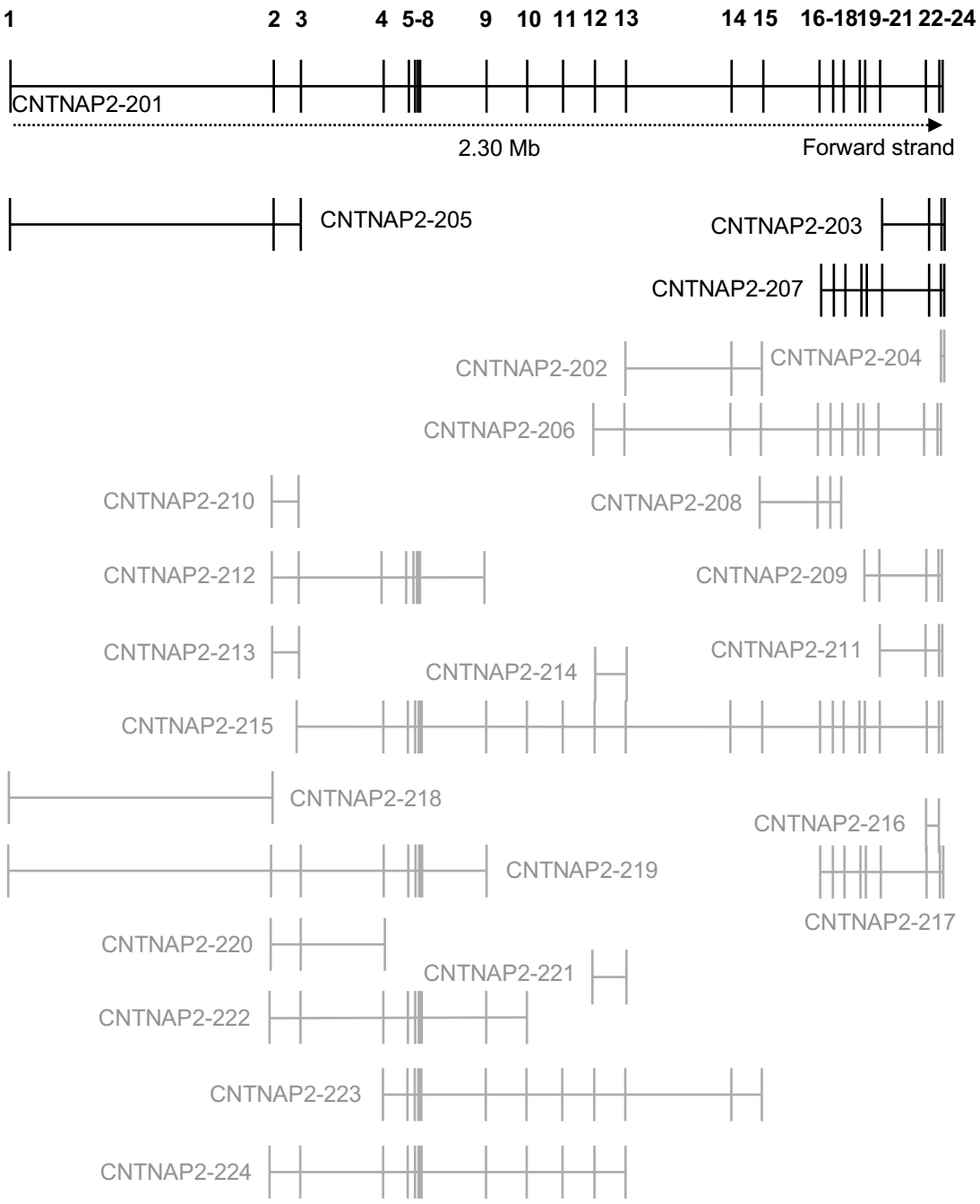


Figure 2.02. Transcripts of *CNTNAP2*

There are 25 transcripts of *CNTNAP2*, only four of which are protein-coding (black: coding transcripts, grey: non-coding). Transcript 201 encodes an mRNA of 9896 nucleotides and 1331 amino acids; 203 encodes 6076 nucleotides and a protein of 108 amino acids; 205 encodes 676 nucleotides and 119 amino acids; and 207 encodes 1944 nucleotides and 390 amino acids. Data taken from Ensembl release 99 – January 2020.

2.2 *CNTNAP2* expression

Although *CNTNAP2* was first described in the myelinated nerves of the peripheral nervous system (PNS) (191), it is primarily expressed in the human central nervous system (CNS) (see **Figure 2.03**). The highest expression is observed in layers II-V of the frontal and temporal cortex, with the strongest expression in layers II-III (195, 196). Outside of the cortex, *CNTNAP2* is also present in the thalamus, amygdala, and striatum. As this pattern of expression is dramatically restricted to the cortico-striato-thalamic circuitry that mediates higher cognitive functions, it supports a potential role for *CNTNAP2* in human cognition (197). **Single nucleus RNA sequencing (snRNA-Seq)** studies of adult middle temporal gyrus (MTG) have shown *CNTNAP2* is robustly expressed in excitatory neurons and interneurons, and minimally expressed in astrocytes, oligodendrocytes, oligodendrocyte progenitor cells (OPCs), and microglia (Allen Brain Atlas, MTG dataset (198)) (**Figure 2.04**). In fact, *CNTNAP2* is more strongly expressed in cortical interneurons than glutamatergic neurons. The strongest expression is seen in parvalbumin-positive (PV^+) interneurons and vasoactive intestinal peptide-positive (VIP^+) interneurons, and lower expression in somatostatin-positive (SST^+) interneurons (199).

In the human fetal brain, *CNTNAP2* is expressed in the frontal and anterior temporal lobes, medial ganglionic eminence, striatum, and dorsal thalamus (**Figure 2.05**) (200, 201). Notably, this anterior cortical enrichment is not observed in rodents. In the developing mouse and rat cortex, *Cntnap2* is broadly expressed throughout the brain and is low or absent in the cortical plate (with highest expression – when present – located posteriorly) (197). Even in adulthood *Cntnap2* is never enriched in the rodent frontal cortex, unlike in humans (197). It is less clear, however, how fetal *CNTNAP2* expression in humans compares to our closest primate relatives, such as chimpanzees and gorillas. This is largely due to restrictions on accessing primate fetal brain tissue.

The sub-cellular expression of CASPR2 in cortical neurons has so far only been studied in rodents (34, 202, 203). In mouse cortical cultures, CASPR2 is expressed in puncta along axons, dendrites, dendritic spines, and cell bodies. Super resolution microscopy has shown CASPR2 is localized to both excitatory and inhibitory synapses, as CASPR2 puncta co-localize with PSD95/vGLUT1 and gephyrin/vGAT respectively (203). However, CASPR2 was found to be more abundant in inhibitory synapses - detected in 61% versus only 45% of excitatory synapses.

With regards to transcription, three transcription factors (TFs) are known to modulate *CNTNAP2* expression: 1) FOXP2; 2) STOX1A; and 3) TCF4 (44). As mentioned in **Chapter 1**, *FOXP2* is one of the best established and widely agreed upon 'brain evolution genes'. Given that *FOXP2* modulates neurite outgrowth and may have contributed to language evolution, it is tempting to speculate part of its effect may have been exerted through *CNTNAP2*. In human fetal cortex, FOXP2 inhibits the transcription of *CNTNAP2*, with their expression patterns appearing in anti-phase (195). Experiments in SH-SY5Y cells also found increased *FOXP2* levels correlated with decreased *CNTNAP2* expression (195). The other TFs, STOX1A and TCF4, inhibit and enhance expression respectively, and all three TFs bind within intron 1 of *CNTNAP2* (44).

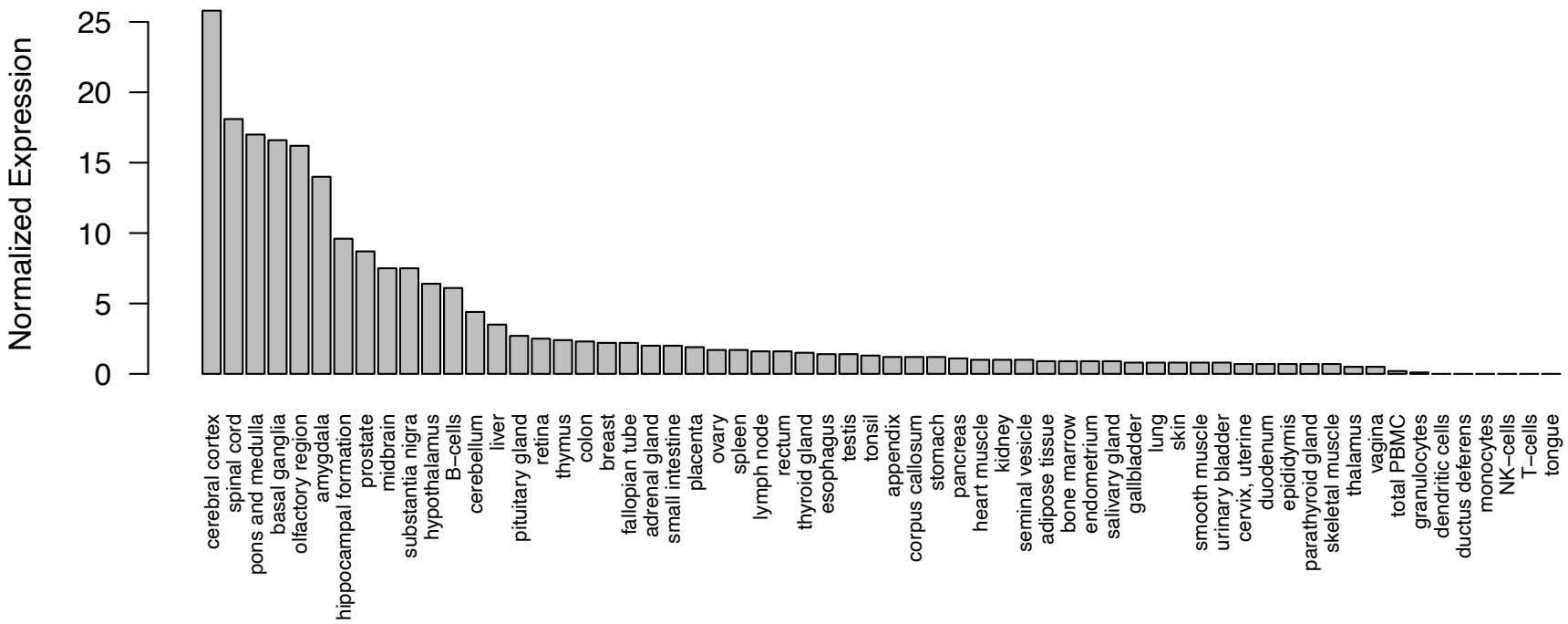


Figure 2.03. Tissue expression of *CNTNAP2*

Normalized RNA expression of *CNTNAP2* in 55 tissue types and 6 blood cell types from the human protein atlas. The profile shown is the consensus of three datasets: RNA-seq from the human protein atlas, RNA-seq from the genotype-tissue expression (GTEx) project, and CAGE data from FANTOM5 project. Highest expression is observed in the CNS, with particular enrichment in the cerebral cortex and spinal cord.

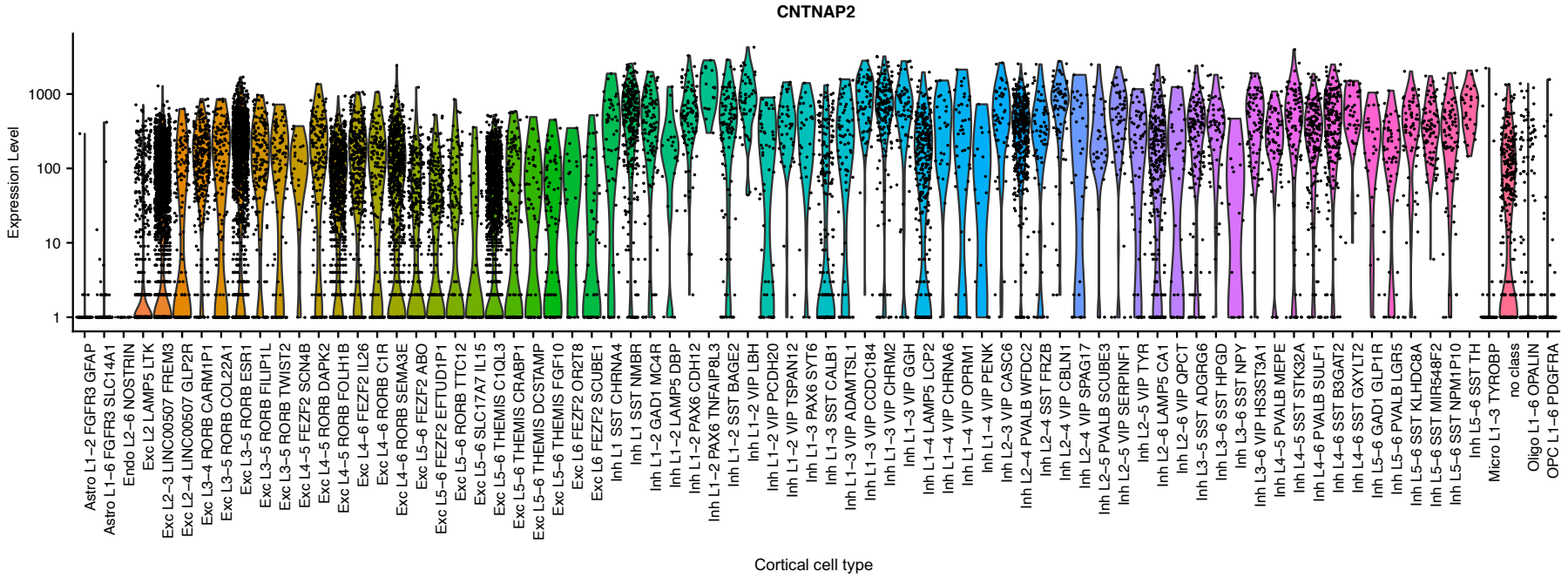


Figure 2.04. Expression of *CNTNAP2* at single cell resolution (adult cortex)

Violin plot of *CNTNAP2* expression in the adult medial temporal gyrus (MTG) (normalized counts, log scale). Data is taken from the Allen Brain Atlas – MTG dataset (Hodge et al. 2019). 75 distinct cell types were identified from 15,928 nuclei (45 inhibitory neuron types, 24 excitatory neuron types, and 6 non-neuronal types). Tissue was derived from from 8 human donors ranging in age from 24-66 years. The highest expression is noted in inhibitory neurons, and the lowest in astrocytes and oligodendrocytes.
Astrocyte (Astro); endothelial cell (Endo); excitatory layer (Exc L-); inhibitory layer (Inh L-); microglia (Micro); oligodendrocyte (Oligo); oligodendrocyte progenitor cell (OPC). Unassigned cells are designated “no class”.

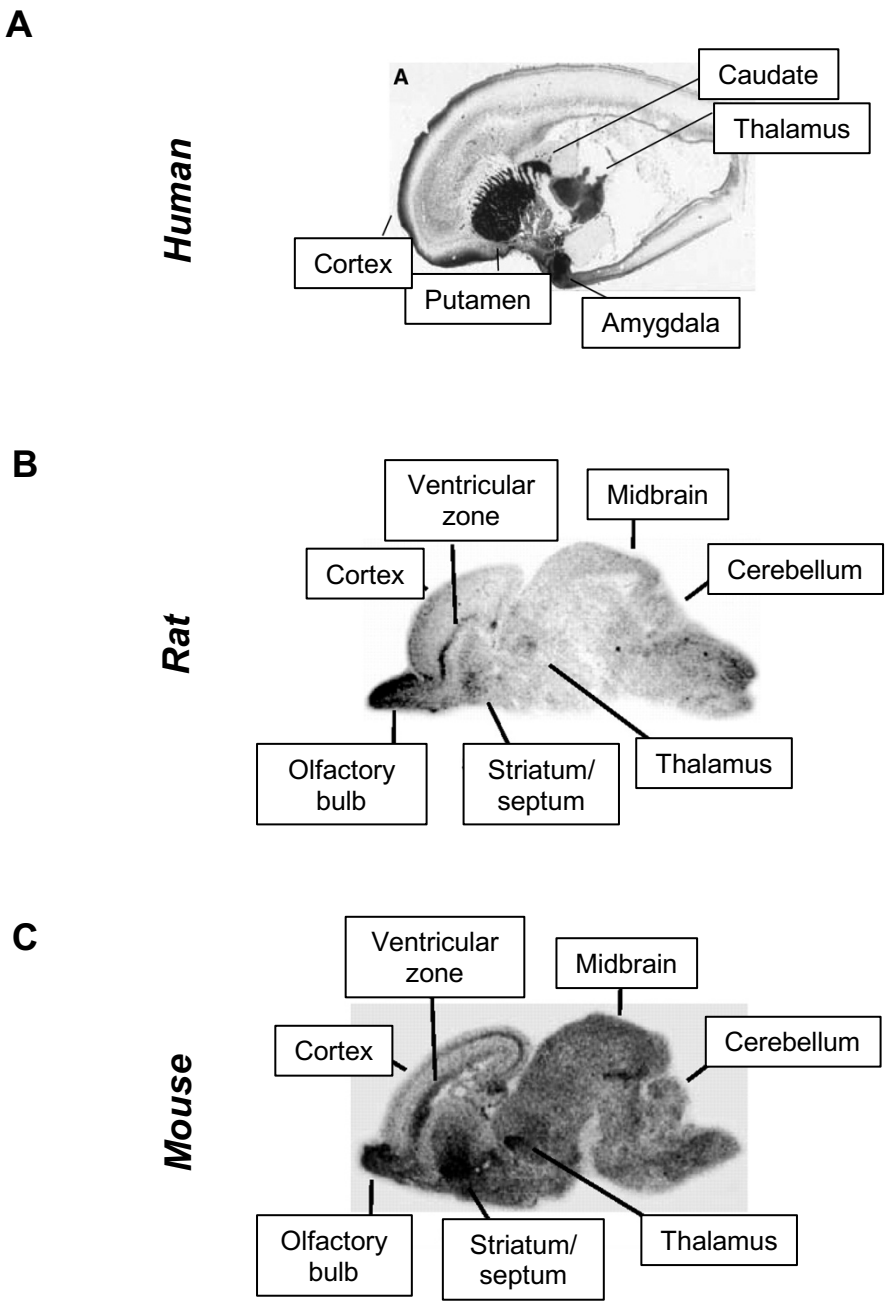


Figure 2.05. Comparison of human and rat fetal *CNTNAP2* expression

(A) Autoradiogram of a midline sagittal section obtained from a human fetal brain (gestational week 18). *CNTNAP2* expression is strongly enriched in the frontal cortex, as well as the thalamus, striatum, and amygdala. Image taken from Alarcón et al. (2008). (B-C) *In situ* hybridizations of *Cntnap2* expression in rat (B) and mouse (C) fetal brain (gestational weeks 20 and 17, sagittal section). Unlike in humans, *Cntnap2* expression in rodents is strongest in the olfactory bulb and low/absent in the cortex. Images taken from Abrahams et al. (2007).

2.2.1 Differential expression amongst primates

In addition to the differences in *CNTNAP2* expression between human and rodent cortex, differential expression between humans and non-human primates has also been observed. Nowick and colleagues (37) found a 1.3-fold increase in *CNTNAP2* expression in the pre-frontal cortex of adult humans compared to chimpanzees. This finding has since been corroborated by Muntane et al. (39), who also showed increased expression on the human-chimpanzee lineage (in comparison to Gibbons, Old and New World Monkeys, Lemurs, and Lorises). Thirdly, Schneider et al. (38) found an almost two-fold increase in the minor transcript, *CNTNAP2-207,* though no significant difference was observed with the canonical transcript. The authors did, however, identify a total of 0.34 Mb of differentially methylated DNA spanning across the *CNTNAP2* locus – approximately 15% of the gene. Differentially methylated regions (DMRs) included intron 2 (decreased in human), exon 17 (increased), and intron 20 (decreased). Of note was that the DMR in intron 2 is ~300 bp away from rs7794745, a common variant associated with ASD (see **Section 2.3**). As such, there are multiple types of evidence suggesting that *CNTNAP2* may be differentially expressed in the frontal cortex of humans and other primates.

It is important to note, however, that other studies have not detected such differences. In particular, a **single cell RNA sequencing (scRNA-Seq)** study by Pollen et al. (3) found no difference in *CNTNAP2* expression between human and macaque primary fetal cortex (post-conception weeks 9-22). In PSC-derived cerebral organoids, the authors also found *CNTNAP2* was more highly expressed in chimpanzee radial glia and excitatory neurons than in human. While these findings may reflect differences in development, maturation and/or *in vitro* artefacts, more work is needed to clarify whether differential expression exists, and if so, in what direction and in what cortical cell types it is. These questions will be discussed further in **Chapter 4**.

2.3 *CNTNAP2* and human disease

Individuals with *CNTNAP2* loss-of-function mutations are typically diagnosed with one of four core conditions: 1) intellectual disability (ID)(204-207), 2) autism spectrum disorder (ASD)(196, 200, 208-224), 3) specific language impairment (SLI) (195, 225-231), and 4) epilepsy (232-234).

As mentioned in **Chapter 1**, ASD is characterized by abnormal language development, restricted interests, stereotyped behaviors, and impaired social skills (196). Autism is a highly heritable disorder - with 70-90% concordance in identical twins (213) - and genetically heterogenous, with multiple genes, pathways and mutations of varying penetrance involved (212). Despite numerous ASD-risk genes being discovered over the past few decades, the genetic basis for many cases remains unknown (213). Similarly, speech and language disorders are believed to have a strong heritable component, though it is generally agreed inheritance is multi-factorial (195, 235). In individuals with *CNTNAP2* mutations, SLI can manifest as dysarthria, language delay, and absent language and/or speech. With respect to epilepsy, it is widely appreciated to have a genetic component, though as with ASD, the precise gene(s) involved are often unknown (232). Epilepsy is one of the most common neurological disorders in humans, with a lifetime incidence of ~3%. In *CNTNAP2* patients, seizures are severe and frequent – they typically begin in early infancy (between 13-36 months), occur several times a day, and are difficult to treat (189). Other disorders associated with *CNTNAP2* include schizophrenia (236-239), attention deficit hyperactivity disorder (ADHD) (240), Tourette syndrome (241), dyslexia (242, 243), and major depression (43). These symptoms have been identified in individuals with microdeletions or point mutations affecting only the *CNTNAP2* locus. As these disorders all target human-specialized behaviors, this supports the notion that *CNTNAP2* may have contributed to the emergence of human cognitive traits.

2.3.1 Pathogenic CNTNAP2 mutations

Most *CNTNAP2* mutations are heterozygous, suggesting both alleles are typically required for normal cognitive function. Crucially however, homozygous mutations cause the most severe phenotypes, and are often found in children of unaffected carrier parents (192). This implies certain *CNTNAP2* mutations are fully penetrant while others are not (see **Table 2.01** and **Figures 2.06-2.07**).

One of the earliest studies of *CNTNAP2* was in an Amish community affected by **cortical dysplasia focal epilepsy (CDFE)** (192). CDFE is characterized by uncontrollable seizures, language regression, social/behavioral disturbances, and intellectual disability (192, 244). All patients were found to be homozygous for a single base deletion in exon 22 (c.3709delG, I1253X) - and their parents found to be healthy carriers. The mutation, which produces a frameshift and pre-mature stop, is believed to cause complete loss-of-function due to the loss of transmembrane and cytoplasmic domains. Post-mortem analyses of patient temporal cortex showed regions of abnormal cortical thickening, increased neuron density, and blurring of the junctions between grey and white matter. The neurons were also abnormally organized into tightly packed clusters, and had a rounded (rather than pyramidal) morphology. Numerous ectopic neurons were identified in sub-cortical white matter and ectopic glia were found in the cortex. *CNTNAP2*-null mice show similar phenotypes: ectopic neurons have been found in the corpus callosum of adult knock out mice, and upper layer neurons were mislocalised to deep layers (layers V-VI) (245).

Aside from CDFE, homozygous mutations in *CNTNAP2* can also cause **Pitt-Hopkins syndrome (PTHS)**. PTHS is characterized by seizures, intellectual disability, hyperventilation, and absence of speech/language (246). Zweier et al. (247) were first to describe two siblings with PTHS caused by homozygous deletions of *CNTNAP2* exons 2-9 (1.15 Mb ; del33-500). They also discovered a second patient with PTHS

that had compound heterozygous mutations within *CNTNAP2;* a deletion of exons 5-8 (180 kb) and a splice site mutation on the second chromosome (IVS10-1G>T). Loss of exons 5-8 was predicted to cause an in-frame deletion, while the IVS10-1G>T mutation was predicted to cause loss of a splice site acceptor leading to a frameshift, deletion of exon 10, and deletion of two laminin-G domains. In both families the parents were heterozygous for the respective mutations but unaffected, again suggesting the loss of a single allele may be tolerated.

Notably, there are clear phenotypic similarities between the patients with CDFE and PTHS. Both disorders cause seizures in early infancy, intellectual disability, and language impairment. More recently described patients with homozygous loss-of-function mutations in *CNTNAP2* also display these phenotypes (189, 205, 207, 248). Lastly, while heterozygous mutations generally cause less severe phenotypes than homozygous mutations, there have been rare exceptions. Gregor et al. (249) described four individuals with heterozygous *CNTNAP2* mutations that cause severe speech/language impairment, epilepsy, and intellectual disability – features usually observed in individuals with biallelic loss of *CNTNAP2*. These mutations, including splice-site, frameshift and pre-mature stop mutations, were either *de novo* or, more curiously, transmitted from a healthy carrier parent. Taken together, these findings suggest there is a large degree of phenotypic variability caused by *CNTNAP2* mutations.

Mutation	*L39X (189)*	*del33-500 (247)*	*A156X (205)*	*S329X (248)*	*E494X (207)*	*del1498-1671 (206)*	*c.1777 + 2T>C (250)*	*C682X (207)*	*I1253X (192, 244)*
Mutation location	Exon 2-3	Exons 2-9	Exon 3-intron 3	Exon 7	Exon 9	Exon 10	Exon 11-intron 11	Exon 13	Exon 22
Mutation size	203 kb	1.15 Mb	76.8 kb	1 bp (c.985delA)	1 bp (c.1480G>T)	173 bp	1 bp	1 bp (c.2046C>A)	1 bp (c.3709delG)
Mutation effect	Frameshift, premature stop	Loss of functional domains	Frameshift, premature stop	Frameshift, premature stop	Premature stop	Loss of exon 10	Splice site disruption	Premature stop	Frameshift, premature stop
No. patients described	2	2	2	1	1	2	1	2	9
Diagnosis	--	PTHS	--	--	--	--	--	--	CDFE
Sex	F, F	M, F	M, F	M	F	M, M	M	M, M	--
Parents	Healthy carriers	Healthy carriers	Healthy carriers	Healthy carriers	Healthy carriers	--	--	Healthy carriers	--
ID	Severe	Severe	Severe	Moderate	Severe	Severe	Severe	Severe	Severe
Speech	No	No	Simple (F), no (M)	No	No	--	--	No	Yes, with regression
Walking	--	Normal	Delayed (4 yrs.)	Delayed (30 mo.)	Delayed (2 yrs.)	--	--	No	Delayed (16-30 mo.)
Age of seizure onset	20-36 mo.	22-30 mo.	2 yrs.	14 mo.	2 yrs.	--	16 mo.	2 yrs.	14-20 mo.
Developmental regression	Yes	--	Yes	Yes, with seizure onset	Yes, with seizure onset	--	--	Yes, with seizure onset	Yes, with seizure onset
Hyperventilation	No	Yes	--	Yes	--	--	--	--	--
Autistic features	Yes	No	--	Yes	Yes	--	--	Yes	Yes (67%)
ADHD	No	--	Yes (M)	--	--	--	--	--	Yes (83%)
Other	Morbidly obese, dysmorphic facial features	Dysmorphic facial features	Small head, ataxic cerebral palsy	Dysmorphic facial features, small head/body	Polydactyly	--	--	Feeding problems	Large head circumference

Table 2.01. Homozygous *CNTNAP2* mutations

Reported homozygous mutations affecting the *CNTNAP2* locus. All patients develop severe seizures in infancy, language impairment, and intellectual disability. Double dashes indicate data was not reported.

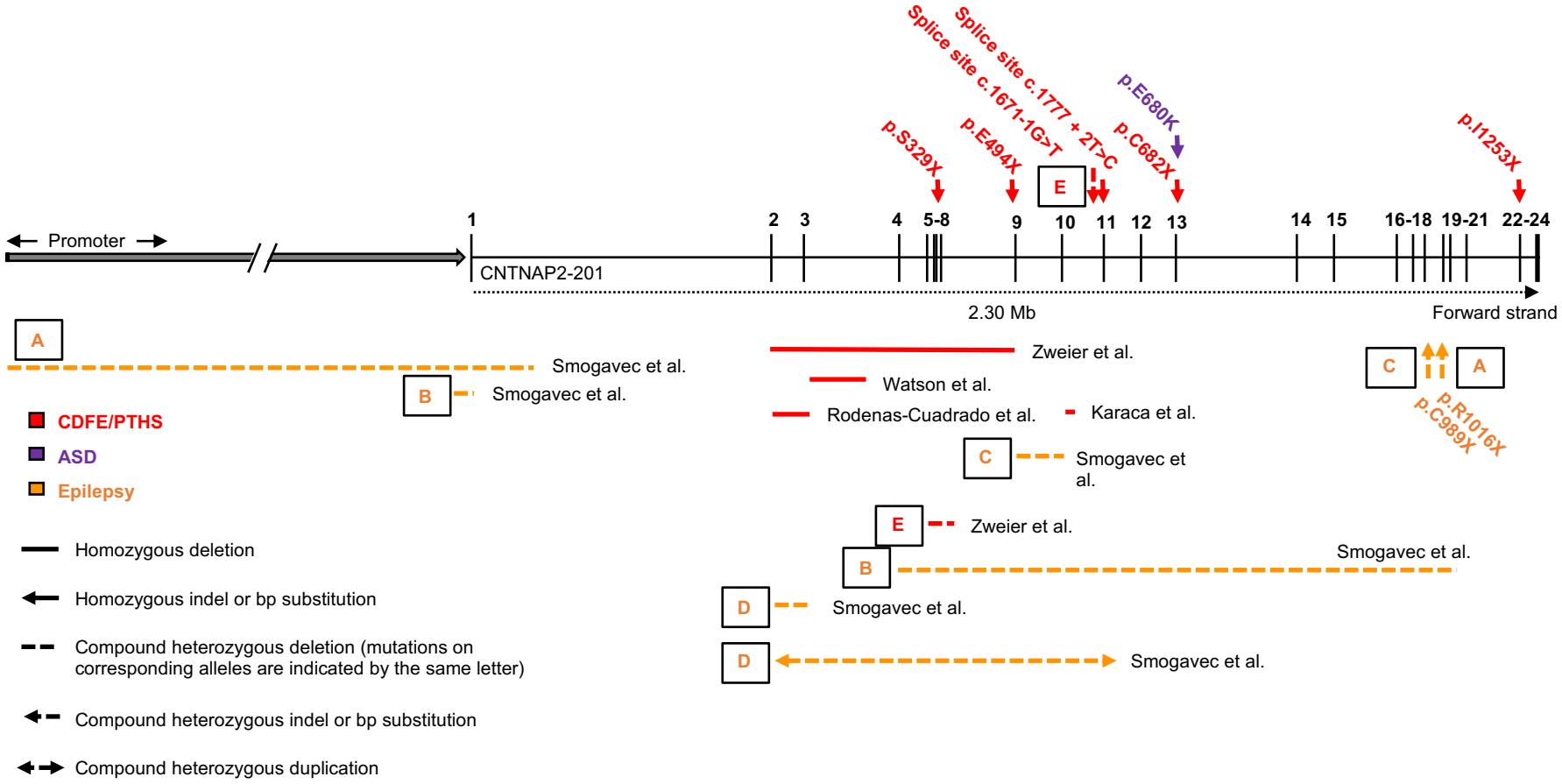


Figure 2.06. Homozygous and compound heterozygous *CNTNAP2* mutations

Mutations affecting the *CNTNAP2* locus are shown with their associated disorders. Mutations are commonly found in the middle of the gene, between exons 2 – 13. CDFE/PTHS (cortical dysplasia focal epilepsy/Pitt-Hopkins syndrome) are the most common diagnoses, though patients typically fall into multiple categories. ASD (autism spectrum disorder). Diagram not perfectly to scale.

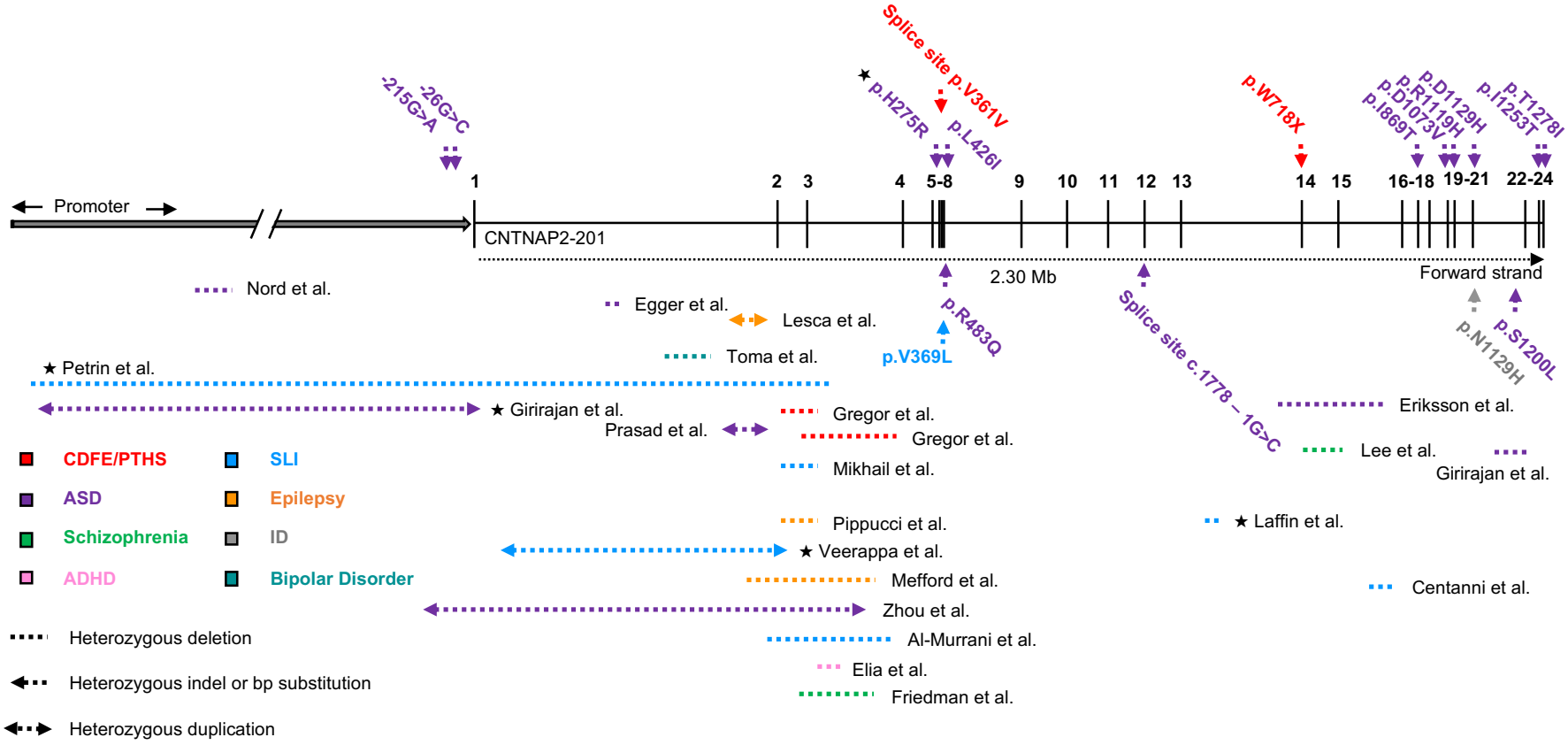


Figure 2.07. Heterozygous *CNTNAP2* mutations

Mutations affecting the *CNTNAP2* locus are shown with their associated disorders. Mutations in patients with multiple affected genes are marked by a star (★). Autism spectrum disorder (ASD) is the most common diagnosis, though patients typically fall into multiple categories. The majority of deletions are found before exon 8. CDFE/PTHS (cortical dysplasia focal epilepsy); ADHD (attention deficit hyperactivity disorder); SLI (specific language impairment); ID (intellectual disability). Diagram not perfectly to scale.

2.3.2 Common CNTNAP2 variants

Common single nucleotide variants (SNVs) within *CNTNAP2* have been implicated in a number of neurodevelopmental disorders (see **Table 2.02** and **Figure 2.08**). Of note are the common variants **rs7794745** (risk allele T) and **rs2710102** (risk allele C), located within intron 2 and 13 respectively. rs7794745 has been repeatedly connected with ASD in a number of family-based association studies (208). Arking and colleagues (208) found that male homozygotes (TT) were significantly more likely to be autistic than females or male carriers. Structural MRIs has shown significant reductions in grey and white matter in the frontal and occipital lobes of TT homozygotes (251). rs2710102 has been connected to ASD (208) and SLI (195, 200), with risk allele carriers showing an association with age of first word, impaired receptive and expressive language skills, and poor performance on tests of nonsense-word repetition (195, 226).

Notably, the majority of disease-related SNVs identified in *CNTNAP2* fall in intron 13. This implies the identified variants are in linkage disequilibrium with the causal variant (and not necessarily that the SNVs are causal themselves). Despite these findings, more recent studies have cast doubt on the significance of *CNTNAP2* common variants to overall disease burden. Sampath et al. (216) assessed 2148 common SNPs across the *CNTNAP2* locus in approximately 2000 families with ASD. No SNP was significantly associated with ASD after correction for multiple testing. Toma et al. (252) came to similar conclusions with ASD as well as schizophrenia and bipolar disorder. In fact, several of the most recent genomic analyses of autism and schizophrenia risk have not identified *CNTNAP2* as a significant contributor to these diseases (see Huckins et al. (253), Skene et al. (254), and Satterstrom et al. (110).

SNV (risk allele)	*Location*	*Risk Allele*	*Associated phenotypes*
rs34712024	Promoter	G	ASD (219)
rs71781329	Promoter	GCG[7] (tandem repeat)	Language impairment (219)
rs7794745	Intron 2	T	Autism (208, 211, 221, 251, 255) Language impairment (211, 226)
rs1603450	Intron 8	G	Language impairment (256)
rs10500171	Intron 8	A	Autism (210)
rs851715	Intron 13	A	Language impairment (195)
rs10246256	Intron 13	T	Language impairment (195)
rs2710102	Intron 13	C	Language impairment (195, 200, 211, 227, 242, 256) Abnormal brain connectivity (255, 257, 258) Depression (259)
rs759178	Intron 13	G	Language impairment (195, 256)
rs1922892	Intron 13	T	Language impairment (195)
rs2538991	Intron 13	G	Language impairment (195)
rs17236239	Intron 13	G	Language impairment (195, 211, 226) Schizophrenia (259)
rs2538976	Intron 13	G	Language impairment (195)
rs2710117	Intron 14	A	Language impairment (195) Depression (259)
rs2710126	Intron 15	A	Schizophrenia (237)

Table 2.02. Disease-associated SNVs within *CNTNAP2*

The most cited *CNTNAP2* common polymorphisms associated with neurodevelopmental disorders. The majority of SNPs cluster in intron 13. (Sources indicated in brackets).

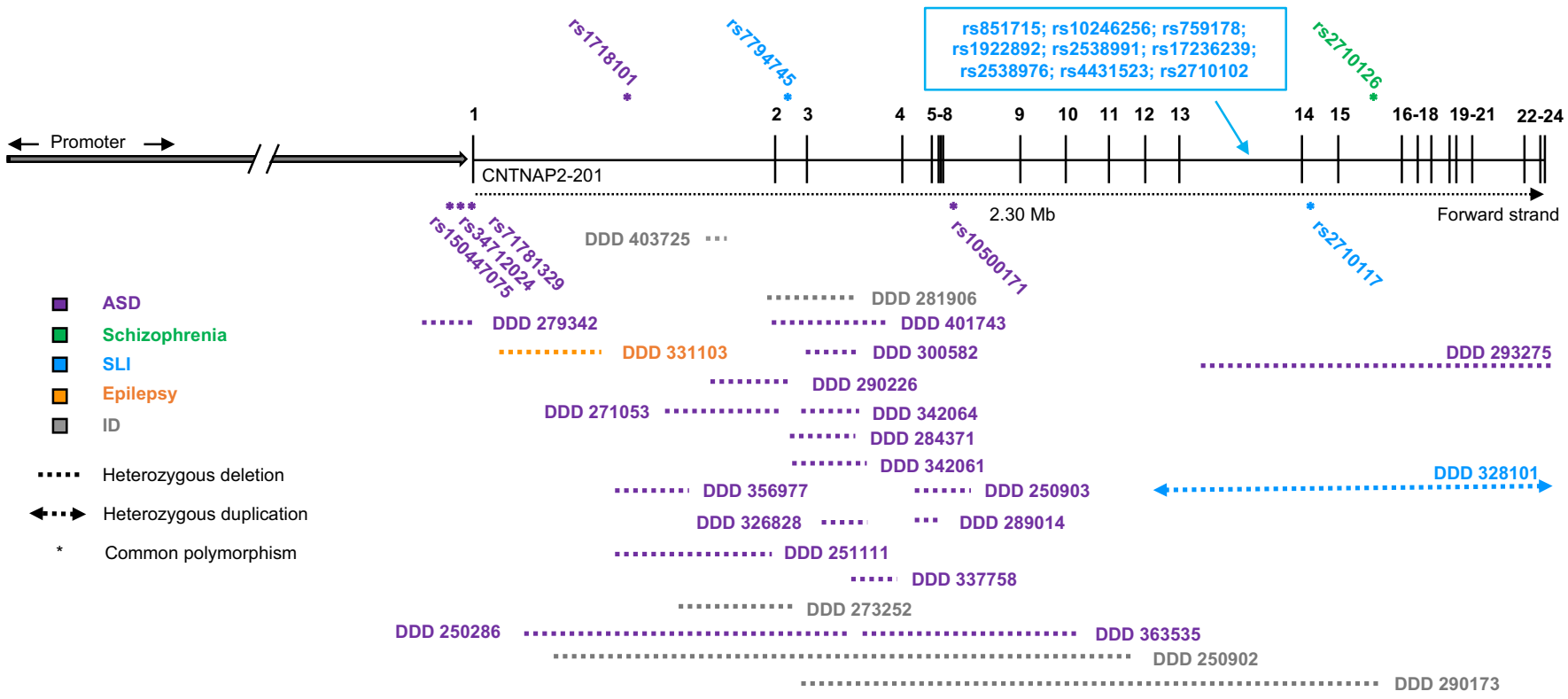


Figure 2.08. Common *CNTNAP2* variants and DDD mutations

Mutations affecting the *CNTNAP2* locus are shown with their associated disorders. Variants identified in the deciphering developmental disorders (DDD) project are denoted by DDD and their identifier code. Mutations in patients with multiple affected genes are marked by a star (★). Autism spectrum disorder (ASD) is the most common diagnosis, though patients typically fall into multiple categories. Many common variants map to intron 13, while may DDD deletions map to introns 2 - 4. CDFE/PTHS (cortical dysplasia focal epilepsy); ADHD (attention deficit hyperactivity disorder); SLI (specific language impairment); ID (intellectual disability). Diagram not perfectly to scale.

2.3.3 Gene dosage

In addition to the association between mutation homozygosity and disease severity, there are other lines of evidence suggesting *CNTNAP2* gene dosage is important for pathogenicity. Specifically, Nord et al. (212) identified an autism patient with a 62 kb deletion in the *CNTNAP2* promoter region. This deletion was confirmed to reduce *CNTNAP2* gene expression in patient-derived lymphoblasts (versus those from healthy controls). Worth noting is that lymphoblasts taken from the patient's healthy mother, who transmitted the mutation to her son, also had reduced *CNTNAP2* expression relative to her non-carrier husband and healthy controls. However, the expression in the affected son was more severely reduced than in the mother.

A separate study from Chiocchetti et al. (219) found a novel variant, g.-215G>A, associated with ASD and delayed age of first word. The variant lies within the *CNTNAP2* promoter and was predicted to disrupt important transcription factor binding sites. In a luciferase assay, the g.-215G>A variant was shown to significantly decrease enhancer potential in SH-SY5Y cells. Other studies suggest not only decreases in *CNTNAP2* expression, but also increases, could be pathogenic. For instance, O'Roak and colleagues (213) described an autistic patient with an inherited missense mutation in *CNTNAP2* (p.His275Arg) and a *de novo* mutation in *FOXP1* (p.Ala339SerfsX4). Using HEK293T cells, the authors discovered the mutant FOXP1 increased the expression of *CNTNAP2* threefold. They hypothesized the increased expression of mutant *CNTNAP2* may have played a causal role in the patient's ASD. Similarly, Lee et al. (238) studied exon-specific expression changes in a schizophrenic patient with a 289 kb deletion in *CNTNAP2* exons 14-15. In patient iPSC-derived neural progenitor cells, *CNTNAP2* expression was significantly decreased in exons 14-15, but unexpectedly, significantly increased in exons 23-24. In a separate follow-up study (239), the authors detected an over 200% increase in spontaneous network level synaptic activity, as well as abnormal expression of several synapse genes.

Intronic disease-causing mutations in *CNTNAP2* have also been discovered: two patients with deletions in intron 1 displayed dysarthric language, autism, and intellectual disability, or ADHD respectively (43). Thus, the expression level of the gene itself is sufficient to cause disorder and may explain the emergence of specific brain phenotypes. Sampath et al. (216) showed expression of *CNTNAP2* in post-mortem cortex of autism patients was significantly altered compared to matched controls. Specifically, both increased and decreased expression were seen, although decreased expression was more common. Finally, snRNA-seq of cortical tissue from autistic patients identified increased *CNTNAP2* expression in layer 2-3 neurons, a cell type believed to be especially vulnerable in ASD (260). Whole exome sequencing revealed 11 of the autistic donors carried heterozygous c.*175dupA or c.*174_*175dupAA mutations in the 3' untranslated region (3'UTR) of the *CNTNAP2* gene. These findings further support that *CNTNAP2* dosage may be important to neurodevelopmental disease pathogenesis – this will be discussed further in **Chapter 4**.

2.4 Role in synaptic function

2.4.1 Neurite length & branching

Dendritic abnormalities have been noted in humans harboring *CNTNAP2* mutations. Post-mortem brain analyses of patients with CDFE (caused by *CNTNAP2* homozygous mutations) have identified neurons in the temporal cortex with irregularly oriented dendritic processes (192). While no additional studies in humans are available, there is strong evidence in mice to suggest *Cntnap2* is involved in neurite development and/or synaptic transmission. In 2012, Anderson et al. (35) showed shRNA-mediated knockdown of CASPR2 in mouse cortical cultures decreased the length and branching of neurites in excitatory neurons. These effects caused a reduction in the amplitude and frequency of excitatory and inhibitory mini post-synaptic currents. Delivery of the CASPR2 shRNA with either lentivirus or calcium phosphate transfection – targeting all neurons or individual neurons respectively – both produced the same results, suggesting the observed effects were cell-autonomous. No changes to dendritic spine density or synapse density were observed, however the width of spine heads was significantly reduced.

A subsequent study by Canali et al. (261) examined neuronal cultures from both homozygous and heterozygous knockout mice. Homozygous knockout neurons had significantly reduced axon lengths, while heterozygous knockout resulted in axons of intermediate length between homozygous knockout and wild type. In addition to a putative role for *CNTNAP2* in excitatory neurite development, Gao et al. (202) proposed *Cntnap2* was involved in neurite growth, but that this effect was specific to cortical interneurons - and is not observed in glutamatergic neurons. The authors showed mature interneurons from *Cntnap2* knockout mice (*in vitro* and *in vivo*) had reduced dendritic branching and dendritic length. No phenotype was observed in excitatory neurons from the same knockout mice or cell cultures. *Cntnap2* overexpression also rescued the phenotypes in interneurons but had no effect in

excitatory cells. Finally, as no difference in branching or length was observed in immature neurons (in either inhibitory or excitatory cells), the authors concluded the reduction in branching was due to decreased neurite stabilization in mature neurons and not impaired outgrowth.

2.4.2 Dendritic spine density

In addition to the proposed role for *Cntnap2* in neurite branching and stabilization, several studies have also suggested the gene may regulate dendritic spine density and/or synapse density. In 2015, Gdalyahu et al. (36) were first to report a reduction in dendritic spine density in an *in vivo* study of Thy1-GFP/*Cntnap2* null mice. Using 2-photon laser scanning microscopy, the authors showed knockout mice had significantly reduced spine density in cortical layer Vb. The reduction was found to be caused by decreased stability of newly formed spines (i.e. loss of spines shortly after they form). No reduction in the formation of new spines was observed, nor was any effect on the maintenance or pruning of already-formed/stable spines. This suggests *CNTNAP2* may be required for the stabilization of new synaptic contacts, a process thought to be critical for the consolidation of adaptive behaviors.

In that same year, Varea et al. (34) reported *in vitro* knockout cultures also had significantly reduced dendritic spine densities. Additionally, the authors observed reduced **GluA1** AMPA receptor subunit expression in spines, and GluA1 cytoplasmic aggregates in cell bodies. These aggregates were found to contain trafficking proteins (e.g. clathrin and rab5), suggesting loss of *Cntnap2* could affect intracellular GluA1 transport. Two other papers reported similar abnormalities in glutamate receptor expression (203, 262). Both papers observed reduced GluA1 expression in knockout cultures and an associated reduction in mEPSC amplitude. As no change to mEPSC frequency was observed, this suggests defects are post-synaptic and not pre-synaptic in nature.

Finally, in 2019 Lazaro and colleagues (263) added further evidence that *Cntnap2* loss-of-function reduces dendritic spine density *in vivo*. The authors reported knockout mice had significantly decreased spine densities and synaptic inputs in layer II-III excitatory neurons. This resulted in reduced amplitude and frequency of mEPSCs in the pre-frontal cortex of knockout mice. *Cntnap2* null mice

also had a reduced density of multi-synapse boutons (a marker of synaptogenesis), and an increase in the number of perforated synapses (a marker of synaptic turnover (36)). These effects led to a two-fold decrease in the frequency and amplitude of mEPSCs. No differences in intrinsic neuronal excitability, neurotransmitter release probability, or synapse maturity were observed between genotypes. Lastly, the study found knockout neurons had reduced network synchrony and less precise firing patterns.

2.4.3 Cortical connectivity

Resting state functional MRIs (rsfMRIs) have shown *Cntnap2* null mice have significantly reduced local and long-range connectivity in the pre-frontal and retrosplenial cortex (264). These reductions were corroborated with rabies virus monosynaptic retrograde tracing, and were believed to underly the defective social behavior observed in the knockout mice (e.g. reduced social exploration and sociability). In an independent study, Zerbi et al. (265) reported similar hypoconnectivity in the pre-frontal cortex of *CNTNAP2* knockout mice. They also observed the defects in connectivity emerge gradually during adolescence and adulthood.

These findings differ somewhat to fMRIs performed on children carrying the common genetic variant rs2710102 (described above) (257). In this study, participants undertook a classification task known to activate cortical regions involved in language processing. Individuals with the 'C' risk allele (heterozygous and homozygous) showed significantly decreased activity in the pre-frontal cortex. However, risk allele carriers also showed a broad pattern of bilateral dorsal and ventral frontal lobe connectivity. Individuals homozygous for the non-risk allele, conversely, showed discrete left-lateralized connectivity that was restricted to inferior frontal cortex and anterior/superior temporal cortex. The authors also observed greater long-range connectivity in the non-risk group (anterior-posterior), and greater local connectivity in the risk group (frontal). These findings are consistent with literature suggesting abnormal cortical asymmetry is involved in autism pathogenesis (266). The findings also agree with data showing increased local connectivity and reduced long-range connectivity in the frontal lobe of autism and SLI patients (267).

Subsequent studies have provided further evidence that variants within *CNTNAP2*, namely rs2710102 (255, 258), rs7794745 (risk allele T) (255), and

rs2710126 (risk allele A) (237) may also cause abnormal cortical connectivity and/or lateralization of language regions. For example, Tan et al. (251) showed variant rs7794745 (risk allele T) reduced grey and white matter in the frontal cortex. These reductions were suggested to be a result of abnormal development or growth of axons.

2.5 *CNTNAP2* in cortical interneurons

CNTNAP2 is robustly expressed in interneurons, and in the ganglionic eminence where interneurons derive from (245, 268). In 2011, Peñagarikano et al. (245) were first to describe a loss of GABAergic interneurons in *Cntnap2* null mice. The authors noted knockout mice had a significant reduction in the number of interneurons in all cortical layers. Parvalbumin positive (PV^+) interneurons were the most affected (20% loss), while calretinin- (CALB2) and neuropeptide Y- (NPY) positive neurons also experienced significant reductions. The loss of interneurons was hypothesized to underly the frequent seizures observed in the mice (as reported by others, see (269, 270)). *In vivo* 2-photon calcium imaging of layer II-III neurons revealed firing was highly asynchronous relative to wild type. The authors did not detect any changes to firing amplitude or frequency, suggesting the asynchronicity was not due to abnormal neuronal activity/conduction, but to defects in synaptic networks.

These findings were supported by further work from Selimbeyoglu et al. (271), who found that the PV^+ interneurons of *Cntnap2* knockout mice had significantly decreased activity *in vivo*. Activating PV^+ interneurons or inhibiting excitatory neurons rescued the observed excitation : inhibition imbalance. Thirdly, Vogt et al. (199) found *Cntnap2* knockout mice had a 24% decrease in PV^+ cortical interneurons, and a 13% decrease in reelin interneurons. No effect was observed on somatostatin (SST) or vasoactive intestinal peptide (VIP) interneurons. Finally, a recent 2020 study by Hali et al. found significant reductions in the number of interneurons in cortical organoids derived from *Cntnap2* knockout mice (272). No differences in glutamatergic neurons were observed. The authors also noted knockout organoids had dramatically reduced expression of transcription factors expressed in ventral telencephalic progenitor cells (e.g. *Dlx2*, *Nkx2.1*, and *Ascl1*). Similar results have also been observed in zebrafish knockout models (269). At 4 days post-fertilization, *Cntnap2* null zebrafish had a 34% reduction in the number of GABAergic neurons in the forebrain. This associated with a significant increase in seizure susceptibility

when the zebrafish were treated with pentylenetetrazol (PTZ), a GABA receptor antagonist known to induce seizures in fish and rodents. Like other *in vivo* mouse models of *Cntnap2* knockout, zebrafish were observed to be hyperactive (245, 270, 273).

2.6 The *CNTNAP2* human accelerated regions (HARs)

CNTNAP2 is unusual for containing eight human accelerated regions:

1. HACNS_116[3] (11);
2. HACNS_97 (11);
3. 2xHAR.395 (10);
4. HACNS_884 (11);
5. ANC1208[4] (13);
6. HACNS_590 (11);
7. ANC1209 (13);
8. HACNS_954 (11)

While this is likely related to the sheer size of the gene (one of the largest in the human genome) (42), it still suggests the locus may be of evolutionary interest. Moreover, the density of HARs is still higher than expected. As mentioned in **Chapter 1,** approximately 3000 HARs have been identified so far (170). Since *CNTNAP2* is 2.3 Mb long, one would expect only two HARs to fall within the gene (assuming that HARs are evenly distributed across the genome). From this perspective, eight HARs would be unexpectedly high.

All of the *CNTNAP2* HARs are found in introns, with half located in intron 1 – a common location for gene regulatory elements (see **Table 2.03** for a summary) (274). The HARs range in length from 23 bp (2xHAR.395) to 510 bp (HACNS_954), and from 3 to 7 human-specific substitutions. Multiple species alignments for each HAR can be found in the **Appendix**. Most of these human-specific changes are shared with Neanderthals and Denisovans, indicating they arose before the emergence of *Homo*

[3] **HACNS:** human accelerated conserved non-coding sequence
[4] **ANC:** accelerated conserved non-coding sequence

sapiens. However, HACNS_97, ANC1209, ANC1209, and HACNS_954 each contain one substitutions that is unique to modern humans alone (166).

HAR	*Intron*	*Coordinates*	*Length (bp)*	*No. of human substitutions*	*No. polymorphic*	*No. shared with archaic humans*	*No. A/T to G/C*
HACNS_116 (Prabhakar)	1	chr7: 146,214,973 – 146,215,168	196	5	1	5	2
HACNS_97 (Prabhakar)	1	chr7: 146,290,445 – 146,290,553	109	5	0	4	2
2xHAR.395 (Linblad-Toh)	1	chr7: 146,420,329 – 146,420,351	23	3	0	3	2
HACNS_884 (Prabhakar)	1	chr7: 146,654,063 – 146,654,409	347	5	0	5	3
ANC1208 (Bird)	11	chr7: 147,516,200 - 147,516,488	289	4	0	3	1
HACNS_590 (Prabhakar)	13	chr7: 147,859,118 – 147,859,418	301	4	0	4	1
ANC1209 (Bird)	13	chr7: 147,878,720 – 147,878,918	199	3	0	2	1
HACNS_954 (Prabhakar)	18	chr7: 148,173,396 – 148,173,905	510	7	0	6	3

Table 2.03. *CNTNAP2* human accelerated regions

Eight human accelerated regions (HARs) lie within the *CNTNAP2* locus. Of these, four are located within intron 1 and two are located in intron 13. The number of human-specific nucleotide substitutions is shown in column 5, followed by the number of these substitutions that are polymorphic in humans (column 6), the number shared in Neanderthals/Denisovans (column 7), and the number of substitutions that are G/C in humans from A/T in other primates (i.e. weak to strong). The majority of *CNTNAP2* HARs appear to be composed of fixed substitutions that are shared with archaic humans. Between 1/4 to 1/2 are weak to strong transitions. Coordinates map to human genome GRCh38; original sources describing each HAR are shown in brackets. Data taken from Hubisz and Pollard (2014).

As mentioned in **Chapter 1**, there are a number of caveats that apply to the interpretation of human accelerated regions including, foremost, (1) the low reproducibility between studies that have used different methods to detect HARs and (2) the dependence of the size of human reference panels in defining which mutations are likely to be fixed in humans species-wide. Each of the *CNTNAP2* HARs was identified by only one of the six major HAR publications (excluding HACNS_884, which was identified by both Gittelman et al. (275) and Prabhakar et al. (11)). Secondly, one of the HARs, HACNS_116, has a human-specific substitution that was subsequently found to be polymorphic (12). All remaining HAR substitutions appear, according to currently available evidence, to be fixed in humans.

Thirdly, GC-biased gene conversion (gBGC) could have caused their signatures of acceleration - and not positive selection (169). Of the eight HARs, between ¼ and ½ of all human-specific changes are [A or T] to [G or C] (i.e. weak to strong). HACNS_97, a 109 bp sequence, has a ~3% increase in GC-content compared to chimpanzees, and 2xHAR.395 (only 23 bp) has a ~4% increase. The remaining HARs have negligible changes to %GC. For context, HAR2, a human accelerated region thought to have undergone gBGC, has a ~9% increase (169). Moreover, the **phastBias tool** (276), an algorithm that predicts gBGC events in the human genome, did not detect gBGC at any of the eight HARs. These are, however, not definitive conclusions about whether gBGC, relaxation of constraint, and/or positive selection have occurred.

That said, evidence for **selective sweeps** has previously been found in *CNTNAP2* introns 1 and 13 (40, 41). A 'selective sweep' occurs when a positively selected variant rapidly increases in frequency, and nearby linked variants rise in frequency along with it as there is no time for recombination to break down the linkage between them (63). Sweeps cause genetic diversity in the region around the selected variant to decrease (168). Strong signals of selection were detected by Ayub et al. (40) in several targets of *FOXP2*, including *CNTNAP2* in some but not all human populations, suggesting selection occurred after the split from Neanderthals and Denisovans (who carry the ancestral sequences) and after the Out-of-Africa dispersal. Introns 1 and 13

each contain several HARs: HACNS_116, HACNS_97, 2xHAR.395, and HACNS_884 in intron 1, and HACNS_590 and ANC1208 in intron 13. However, only HACNS_97 coincides directly with one of the sweeps.

There is also preliminary evidence to suggest one or more of the *CNTNAP2* HARs could be enhancers. HACNS_884 was shown by Gittelman et al. (275) to overlap a human-specific **DNase I hypersensitive site (DHS)**. DNase I selectively cleaves regions of open/active DNA, which is the expected chromatin state of regulatory elements (277). Importantly, the presence of a DHS does not mean a sequence is functional or phenotypically significant. It does, however, help prioritize potential regulatory loci for further testing. Won et al. (175) further identified six of the eight HARs as overlapping DHSs in fetal brain (all except 2xHAR.395 and ANC1208). Finally, Capra et al. (170) detected HACNS_884 and HACNS_954 as putative enhancers using their enhancer finding pipeline (discussed in **Chapter 1**). They were also able to bioinformatically predict that HACNS_884 is active in fetal brain, but could not provide a functional prediction for the role of HACNS_954. Interestingly, HACNS_884 also contains a FOXP2 transcription factor binding site, thereby providing yet another potential link between the *CNTNAP2* HARs, gene regulation, and human evolution (278).

Finally, as mentioned in **Chapter 1**, HARs appear to be enriched in human disease-causing mutations. This is also true for the *CNTNAP2* HARs which overlap mutations associated with autism (ASD), specific language impairment (SLI), epilepsy, and/or intellectual disability (ID) (**Figure 2.09**). Curiously, all of these mutations are heterozygous (or compound heterozygous), which suggests that the loss of even a single copy of a HAR may be enough to cause disease. However, this relies on the assumption that these mutations are pathogenic due to the loss of the HAR. This cannot be assumed, particularly because many of the mutations are large, and therefore not specific. Moreover, most of the mutations encompass multiple HARs, suggesting it could be a combinatorial loss of HAR function(s) that leads to disease.

HACNS_884 overlaps the most mutations (11 in number), followed by HACNS_116 (6 in number), and 2xHAR.395 with five mutations.

Taken altogether, there is preliminary data that one or more of the *CNTNAP2* HARs are functional and may be important for human cognition. As with many loci uncovered by genome-wide analyses, further experimental studies are needed to clarify the specific roles of these HARs (if any), and the precise mechanisms by which they work. **Chapter 7** will combine evolutionary genetics with wet-lab functional experiments to shed light on these and other questions.

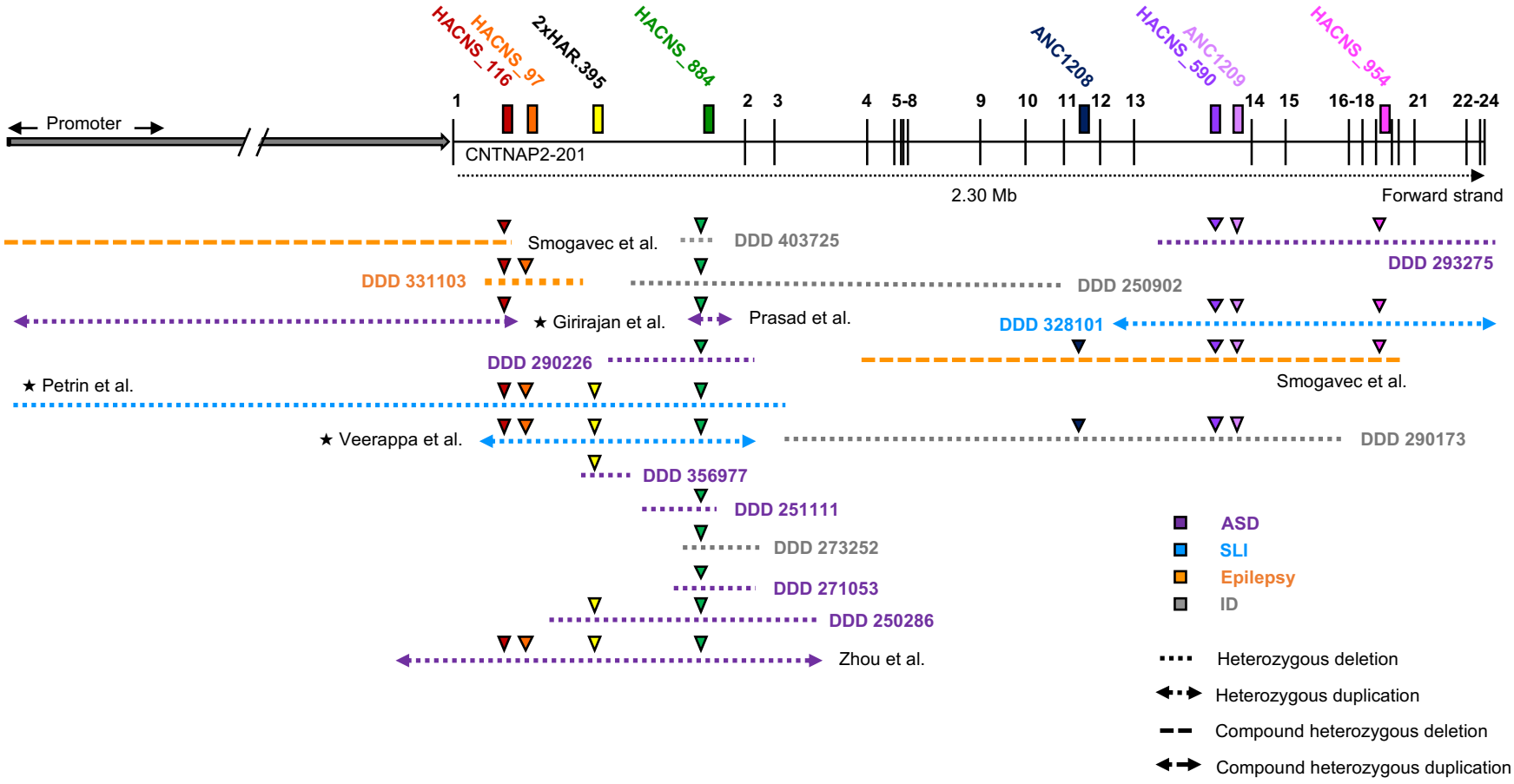


Figure 2.09. Overlap of human accelerated regions (HARs) and *CNTNAP2* mutations

Disease-causing mutations that overlap one or more human accelerated region. Coloured triangles indicate overlap: the colour represents the specific HAR involved. Mutations in patients with multiple affected genes are marked by a star (★). CNS884 in intron 1 intersects the most mutations (11), while ANC1208 in intron 11 intersects the fewest (2). ASD (autism spectrum disorder); SLI (specific language impairment); ID (intellectual disability); DDD (deciphering developmental disorders). Diagram not perfectly to scale.

Chapter 3

Materials and methods

3.1 Materials

3.1.1 Cell lines

The following cell lines were used in this project:

Homo sapiens

1. H9 embryonic stem cells (ESCs) (Wicell Research Institute)
2. SFC840 induced pluripotent stem cells (iPSCs) (StemBANCC)
3. AD2.1 iPSCs (Israel et al. (279))
4. NDC1.2 iPSCs (Israel et al. (279))
5. HEK-293T human embryonic kidney line
6. SH-SY5Y neuroblastoma line

Macaca fascicularis

1. MF1 ESCs (Washington National Primate Research Centre, WaNPRC)
2. MF12 ESCs (WaNPRC)

3.1.2 Human PSC reagents

Product name	Supplier	Catalogue number
Geltrex	Life Technologies	A1569601
DMEM:F12 + glutamax	Life Technologies	31331
Rock inhibitor (Y-27632)	Tocris Bioscience	1254
Essential-8 basal media and supplement	Life Technologies	A1517001
0.5M EDTA	Life Technologies	15575020
DPBS	ThermoFisher	14190094
DMSO	Sigma	D2650

hPSC freezing media

1. Essential-8 media
2. 10% DMSO
3. 10μM rock inhibitor

3.1.3 Macaque PSC reagents

Product name	Supplier	Catalogue number
KnockOut serum replacement	Gibco	10828028
Essential-8 basal media and supplement	Life Technologies	A1517001
Pens/Strep (10000 U/ul)	Life Technologies	15140
Wnt antagonist I, IWR-1-endo	Sigma	681669
DMEM + glutamax	Life Technologies	61965059
Rock inhibitor (Y-27632)	Tocris Bioscience	1254
DMSO	Sigma	D2650
Fetal bovine serum (FBS)	Life Technologies	16000-044
Mouse embryonic fibroblasts	AMS Biotechnology	GSC-6001G
Dispase	Life Technologies	17105041

TOM media

1. Essential-8 media
2. 5% KnockOut serum replacement
3. Pen-strep
4. 1:1200 IWR-1

STO media

1. DMEM + glutamax
2. 10% filtered FBS

mPSC freezing media

1. TOM media
2. 10% DMSO
3. 10μM rock inhibitor

3.1.4 Cortical induction reagents

Product name	Supplier	Catalogue number
Neurobasal media	Life Technologies	12348
DMEM:F12 + glutamax	Life Technologies	31331
N2 supplement	Life Technologies	17502048
B27 supplement	Life Technologies	17504044
Insulin (10mg/ml)	Sigma	I9278
2-mercaptoethanol (50mM)	Life Technologies	31350
Non-essential amino acids (100x)	Life Technologies	11140
Sodium Pyruvate (100mM)	Sigma	S8636
Pens/Strep (10000 U/ul)	Life Technologies	15140
Glutamax (200mM)	Life Technologies	35050-038
SB431542	Tocris	1614
Dorsomorphin	Tocris	3093
FGF2	Peprotech EC Ltd	100-18B-50
Laminin	Sigma	L2020
Dispase	Life Technologies	17105041
Accutase	Sigma	A6964-100ml
DMSO	Sigma	D2650

Neural maintenance media (NMM)

1. DMEM:F12 + glutamax
2. Neurobasal
3. 5µg/mL insulin
4. 100µM 2-mercaptoethanol
5. 100µM non-essential amino acids
6. 1mM Sodium Pyruvate
7. 1mM glutamax
8. PenStrep
9. N2 supplement
10. B27 supplement

Neural induction media (NIM)

1. NMM
2. 10µM SB431542
3. 1µM dorsomorphin

Neural freezing media (NFM)

1. NMM
2. 10% DMSO
3. 20ng/ml FGF2

3.1.5 SH-SY5Y reagents

Product name	Supplier	Catalogue number
DMEM + glutamax	Life Technologies	61965059
Fetal bovine serum (FBS)	Life Technologies	16000-044
TrypLE express enzyme	Life Technologies	12604-013
DMSO	Sigma	D2650

SH-SY5Y freezing media

1. FBS
2. 10% DMSO

3.1.6 Tissue culture materials

Product name	Supplier	Catalogue number
6-well cell culture plate	ThermoFisher	140675
10cm cell culture dish	Sigma	CLS430167
24-well cell culture plate	ThermoFisher	142475
Nunc EasYFlask 75cm^2 with filter	ThermoFisher	156499
96-well imaging plate	ibidi	89626
1.5mL cryovials	Greiner	121279
Coolcell LX	Corning	432002

3.1.7 Transfection/transduction materials

Product name	**Supplier**	**Catalogue number**
P3 primary cell 4D-nucleofector X kit S	Lonza	V4XP-3032
P3 primary cell 4D-nucleofector X kit L	Lonza	V4XP-3012
4D nucleofector core unit	Lonza	AAF-1002B
4D nucleofector X unit	Lonza	AAF-1002X
FuGENE HD	Promega	E2311

3.1.8 Plasmids

Product name	**Supplier**	**Catalogue number**
pSynapsin1-mNeonGreen	Generated by R. Solanki, Livesey lab	--
pCamkii-mKate2	Addgene	96942
pGL3 control	Promega	E1741
pGL3 promoter	Promega	E1761
pRL	Promega	E2231

3.1.9 Primers and oligonucleotides

Name	Sequence 5' to 3'	Supplier
CRISPR_exon3_forward	AGAGCACTGCCAAGACCAAT	IDT
CRISPR_exon3_reverse	ACCATTGGATAGAAATTACAGCCTG	IDT
Miseq_exon3_forward	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTCATCTTACCTCTGCCCATCTTC	IDT
Miseq_exon3_reverse	TCGGCATTCCTGCTGAACCGCTCTTCCGAT CTTGCCAATGACTTACCCAGATATT	IDT
Miseq_forward4	AATGATACGGCGACCACCGAGATCTACAC GGCTCTGAACACTCTTTCCCTACACGACGC TCTTCCGAT*C*T	M. Haneklaus, Livesey lab
Miseq_reverse	CAAGCAGAAGACGGCATACGAGATAACGT GATGAGATCGGTCTCGGCATTCCTGCTGA ACCGCTCTTCCGATC	M. Haneklaus, Livesey lab
CRISPR_PAC1_ssODN	AGAGCATCCGGTATTGGGTCACCCAATCT GAGCTGCTATACCTTCCTTGGttaattaaGTT GCAATGGCACTGATCTGCTTCCGATTGCCA AAGTCAACCTGAAGCCA	IDT
CNTNAP2_qPCR_forward	TGCAACCCAAGGAAGGTATAG	IDT
CNTNAP2_qPCR_reverse	CCGGGAAATGCCCAGATATT	IDT
CNTNAP2_qPCR_for2	TGTCCTTCAGCCTTCATTCCA	IDT
CNTNAP2_qPCR_rev2	CATGTTTGCGAGCACTTTCCA	IDT
CNTNAP2_qPCR_for3	GGTCACATCTGGGTTCATATCC	IDT
CNTNAP2_qPCR_rev3	GGAGTAACCGTGGTATCTCTCTA	IDT
RAB7A_qPCR_forward	GCCACAATAGGAGCTGACTT	IDT
RAB7A_qPCR_reverse	GTCCTGCTGTGTCCCATATC	IDT
C1orf43_qPCR_forward	TATGAAACAGCCCGCTATGG	IDT
C1orf43_qPCR_reverse	GCTCCCAATTCGTGCTTTAAC	IDT
DDB1_qPCR_forward	GGTGACTCCCAGCTTGTGAA	A. Strano, Livesey lab

DDB1_qPCR_reverse	AGGTCCACCACGCACATATC	A. Strano, Livesey lab
PRPF6_qPCR_forward	GGTTCGCTTGTGGAAAGCAG	A. Strano, Livesey lab
PRPF6_qPCR_reverse	GCCTTGTTCAAGACCTTGCG	A. Strano, Livesey lab
RV3_sequencing	CTAGCAAAATAGGCTGTCCC	IDT
GL2_sequencing	CTTTATGTTTTTGGCGTCTTCC	IDT
CNS97_infusion_forward	CTAGCCCGGGCTCGAGGCCCCTTGTTAAGCTGGTTC	IDT
CNS97_infusion_reverse	TGCAGATCGCAGATCTAACTGACTTCAAAGATGGCGAT	IDT
CNS116_ infusion_forward	CTAGCCCGGGCTCGAGGCTTTGCCAATCCTTCCGAA	IDT
CNS116_ infusion_reverse	TGCAGATCGCAGATCTCAGTTGGCCATTGGATTATTTGAG	IDT
HAR.395_infusion_forward	CTAGCCCGGGCTCGAGTGGAATCTCACCTTCAAGCTTG	IDT
HAR.395_infusion_reverse	TGCAGATCGCAGATCTTGCACCTGTCTCCAATGATCT	IDT
CNS590_infusion_forward	CTAGCCCGGGCTCGAGAATGTCCTCCCGCTTCTCAA	IDT
CNS590_infusion_reverse	TGCAGATCGCAGATCTACTACCTTCTTGTGAGACAGCT	IDT
CNS884_infusion_forward	CTAGCCCGGGCTCGAGCCATAGAAACTGGGCTTGGT	IDT
CNS884_infusion_reverse	TGCAGATCGCAGATCTCATTTATTTCTGGGAAGCCAAGC	IDT
CNS954_infusion_forward	CTAGCCCGGGCTCGAGTTTTCTGTGAGCGGTTGGTG	IDT
CNS954_infusion_reverse	TGCAGATCGCAGATCTCTCCAGAGGTTTTGCGTGAG	IDT

3.1.10 CRISPR-cas9 reagents

Product name	Supplier	Catalogue number
IDTE (1X TE Solution) pH 7.5	IDT	11-05-01-05
P3 Solution	Lonza	V4XP-3012
Nuclease free duplex buffer	IDT	11-01-03-01
tracrRNA	IDT	1072532
SpCas9	Generated by L. Evans, Livesey lab	--
CloneR	StemCell Technologies	05888
T7 endonuclease I	New England BioLabs	M0302S
PAC1 restriction enzyme	New England BioLabs	R0547S

Nuclease free duplex buffer

1. 30 mM HEPES, pH 7.5
2. 100 mM potassium acetate

Cloning media

1. Essential-8 media
2. 10% CloneR

2.1.11 CRISPR-cas9 crRNAs

crRNA name	Sequence 5' to 3'	Supplier
*CNTNAP2*_KO_exon3	ATCAGTGCCATTGCAACCCA	IDT

3.1.12 Molecular biology reagents

Product name	Supplier	Catalogue number
Buffer RLT	Qiagen	79216
Buffer ATL	Qiagen	19076
Qubit RNA BR assay kit	Invitrogen	Q10211
Superscript III reverse transcriptase	ThermoFisher	18080044
Random hexamers	ThermoFisher	N8080127
RNase-free DNase set	Qiagen	79254
PowerUp SYBR green master mix	Applied Biosystems	A25918
Phusion high-fidelity DNA polymerase	New England Biolabs	M0530S
KAPA HiFi HotStart ReadyMix	KapaBiosystems	KR0370
In-Fusion HD cloning plus	Takara Bio	638920
QuickExtract DNA extraction solution	Lucigen	QE09050
Triton X100	Sigma	X100
PFA	Sigma	P6148
Sucrose	Sigma	S7903
Glacial acetic acid	Sigma	A6283
Normal goat serum	Sigma	NS02L
Pierce BCA protein assay kit	ThermoFisher	23225
RIPA Buffer	Sigma	R0278
Protease inhibitor tablets	Sigma	4693116001
NuPAGE LDS sample buffer (4X)	ThermoFisher	NP0007
1M Dithiothreitol (DTT)	ThermoFisher	R0861
Immobilon-P PVDF membrane	Millipore	IPVH00010
NuPAGE MES SDS running buffer (20X)	ThermoFisher	NP0002
NuPAGE 4-12% Bis-Tris protein gels	Invitrogen	NP0335BOX
Tween 20	Sigma	P1379
Pierce ECL western blotting substrate	ThermoFisher	32106

Glycine	Sigma	G7126-1KG
Tris base	Sigma	77-86-1
Methanol	Sigma	34860-1L-R

RIPA buffer solution

1. RIPA buffer
2. DNase I
3. Protein inhibitor tablet

1X transfer buffer

1. 28.8g Glycine
2. 4.06g Tris base
3. 200mL Methanol
4. 1.6L ddH_2O

3.1.13 Luciferase enhancer assay reagents

Product name	**Supplier**	**Catalogue number**
XhoI restriction enzyme	New England BioLabs	R0146S
Dual luciferase reporter assay	Promega	E1910
Opti-MEM I Reduced Serum Medium	Gibco	31985062
Hidex sense microplate reader	Hidex	425-301

3.1.14 Antibodies

Product name	**Supplier**	**Catalogue number**
Anti-CASPR2 (WB & IF 1:1000)	NeuroMab	K67/25
Anti-MAP2 (IF 1:1000)	Synaptic systems	188 004
Anti-TAU (WB 1:5000, IF 1:1000)	Dako Cytomation	A0024
Anti-TUBB3 (WB 1:5000, IF 1:1000)	Biolegend	802001
Anti-ß-ACTIN (WB 1:5000)	Abcam	ab6276
Anti-PSD-95 (IF 1:100)	Abcam	ab18258
Anti-SYNAPTOPHYSIN (IF 1:100)	Abcam	ab14692
DAPI (IF 1:2000)	Sigma	D9542
Alexa fluor secondary antibodies (1:1000)	Invitrogen	--

PFA + sucrose solution

1. 4% PFA
2. 4% sucrose
3. 1X PBS
4. Adjusted to pH 7.5 with 1N NaOH and glacial acetic acid

3.1.15 Single cell RNA sequencing (scRNA-Seq) reagents

Product name	**Supplier**	**Catalogue number**
Accutase	Sigma	A6964-100ml
7.5% Bovine Albumin Fraction V	ThermoFisher	15260037
Actinomycin D	Sigma	A1410-2MG
Papain dissociation system	Worthington	LK003150
TotalSeq-A0251 anti-human Hashtag 1 Antibody	BioLegend	394601
TotalSeq-A0252 anti-human Hashtag 2 Antibody	BioLegend	394603
TotalSeq-A0253 anti-human Hashtag 3 Antibody	BioLegend	394605
TotalSeq-A0254 anti-human Hashtag 4 Antibody	BioLegend	394607
2ml, DNA Lo-Bind tubes	Eppendorf	0030108078
SPRIselect Reagent Kit	Beckman Coulter	B23317
FcR Block	Miltenyi Biotec	120-000-442
Cell Staining Buffer	BioLegend	420201
Chromium Next GEM Single Cell 3′ Kit v3.1	10X Genomics	1000269
Chromium Controller & Next GEM Accessory Kit	10X Genomics	1000202

Dissociation mix

1. 290 µL Papain
2. 20 µL DNase I
3. 90 µL fresh Accutase
4. 1.6 uL Actinomycin D

Inhibitor mix

1. 296.5 µl EBSS
2. 17.5 µl DNase I
3. 35 µl Ovomucoid
4. 1.4 uL Actinomycin D

3.2 Methods

3.2.1 Human PSC culture

All cell lines were grown in a humidified incubator at 37°C with 5% CO_2. Cells were handled under sterile conditions in a class II biological safety hood with laminar flow. hPSC lines underwent copy number variation analysis (CNV) to verify genome integrity and overall quality for experiments.

hPSCs were stored in liquid nitrogen until use. Several hours before thawing, tissue culture plates were coated with geltrex basement membrane matrix. Cells were partially thawed in a 37°C water bath, washed in DMEM:F12 + glutamax, and centrifuged at 400 r.c.f. for 4 minutes. hPSCs were then re-suspended in warm essential-8 ("E8") media supplemented with 10μM rock inhibitor to promote cell survival. Geltrex was aspirated from prepared culture plate, and cells/media immediately plated. E8 was thereafter replaced every 24 hours without rock inhibitor. Once hPSCs reached 70-80% confluency, they were passaged onto a new geltrex-coated plate. To do so, hPSCs were washed with 1X PBS and dissociated using 0.5M EDTA diluted 1:1000 in DPBS. Cells normally detached after 3-4 minutes at room temperature. To freeze hPSCs, the cells were washed with 1X PBS, dissociated with diluted EDTA, re-suspended in hPSC freezing media, and transferred to labelled cryovials. Vials were initially kept in a CoolCell at -80°C overnight, and then moved to liquid nitrogen for long term storage the following morning.

3.2.2 Macaque PSC culture

Macaque PSCs were grown on mouse embryonic fibroblasts (MEFs). One day before mPSC plating, a cell culture dish was coated in 0.1% gelatine at 37°C for 5 minutes. MEFs were partially thawed in a 37°C water bath, washed in STO media, and centrifuged at 400 r.c.f. for 4 minutes. Gelatin was then aspirated from the plate, and

MEFs re-suspended in STO before immediately plating. Thereafter cells were fed every 24 hours with fresh TOM media until confluent. To passage mPSCs, cells were dissociated from MEFs with dispase treatment. Dispase was added directly to the culture media at 1mg/mL, and the cells incubated at 37°C for 20-40 minutes. Once mPSCs were completed detached from feeder cells, they were washed twice in 1X PBS, then re-suspended in TOM media, and transferred to a new MEF-coated plate. To freeze mPSCs, the cells were dissociated with dispase, washed twice, and re-suspended in mPSC freezing media. Cells were then transferred to prepared cryovials and frozen as described for hPSCs.

3.2.3 Cortical differentiation

PSCs were differentiated according to the protocol described by Shi et al. (2012). Once cells reached 70-80% confluency, they were washed with 1X PBS, dissociated with diluted EDTA, and passaged 3 wells into 2 wells with E8 + rock inhibitor. The following day, cultures were checked for complete confluency under a tissue culture microscope. Cells were then washed with 1X PBS and fed with neural induction media (NIM). NIM was replaced daily for the next 11 days while the sheet of cells acquired neuroepithelial identity. On day 12, the neuroepithelial cells were dissociated with dispase treatment. Dispase was added directly to the culture media at 1mg/mL. Cells were incubated at 37°C/5% CO_2 for 20-40 minutes until the sheet completely detached from the plate. The sheet was then washed three times with 1X PBS, manually fragmented, and re-plated 1:2 onto a laminin plate containing NIM.

Between days 13-16 media was replaced daily with neural maintenance media (NMM) supplemented with 20ng FGF2. From day 17-24 the cells were fed every 48 hours with NMM and passaged with dispase when needed. On day 25, cortical rosettes were dissociated into single neural progenitor cells (NPCs). Rosettes were incubated in accutase at 37°C for 7-10 minutes, single-celled by pipetting, washed with 1X PBS, and centrifuged at 400 r.c.f. for 4 minutes. Cells were then transferred to a laminin-

coated plate containing NMM. From day 26 until NPCs were frozen on day 30 or final plated on day 35, NMM was changed every 48 hours. For freezing, NPC cultures were dissociated with accutase, re-suspended in neural freezing media, and frozen as with hPSCs. For final plating, cells were dissociated with accutase, washed/centrifuged, re-suspended in NMM, and plated onto a geltrex plate at a density of 200,000 cells/cm^2. Dissociated cells were counted using LUNA-FL Automated Fluorescence Cell Counter (Logos Bio, L20001).

3.2.4 Nanostring - induction quality control

To verify PSCs correctly differentiated to cortical neurons, RNA from every induction was run on a Nanostring nCounter SPRINT Profiler. In this platform, mRNA from user-defined genes of interest are tagged with barcoded 'reporter probes' (one barcode per gene). RNA was collected from cultures between days 30-35. Cells were dissociated with accutase and centrifuged at 400 r.c.f. for 4 minutes. The resulting supernatant was aspirated, and the cell pellet was re-suspended in lysis buffer RLT. Lysate was either stored at -80°C or immediately processed. RNA was extracted with the RNeasy mini kit according to the manufacturer's protocol (Qiagen, 74106). Purified RNA was stored at -80°C until use. Approximately 16 hours before running on the Nanostring, RNA concentration was measured with a Qubit RNA BR assay. 10ng/μL RNA was then hybridized with the Nanostring reporter and capture probes at 65°C for 16 hours. The following morning, the hybridized samples were loaded onto a Nanostring cartridge and run on the nCounter SPRINT Profiler. The complete list of genes measured by the Nanostring is reported in **appendix 1**. In brief, inductions were considered cortical if they showed robust expression of cortex/forebrain associated genes, and weakly expressed markers of non-cortical cell types (e.g. hindbrain, midbrain, pluripotency factors, etc.).

3.2.5 Quantitative real-time polymerase chain reaction (qRT-PCR)

cDNA was synthesized from 100-500ng total RNA using random hexamers and Superscript III reverse transcriptase according to the manufacturer's protocol. All qPCR primers were designed with Primer3 software, and wherever possible spanned an exon-exon boundary. qPCR reactions were carried out with PowerUp SYBR Green Master Mix and run on a StepOnePlus Real-Time PCR System (Applied Biosystems, 4376600). Primer efficiency was tested through a standard curve of serially diluted cDNA. Primers with 90-110% efficiency, a slope of approximately -3.3, and a $R^2 > 0.99$ were accepted for use (see **appendix 2**). Primer specificity was also verified with Primer-Blast, analyzing melt curve profiles, and by running the qPCR products on a gel. Housekeeping/normalization genes were chosen by consulting Cortecon and Eisenberg & Levanon (2013). For inter-species qPCR assays, primer pairs were validated to have comparable efficiency in all species, and specifically housekeepers (designed by A. Strano, Livesey lab) had comparable expression.

3.2.6 Western Blots

To extract protein, cell pellets were lysed in RIPA buffer solution using a Precellys Evolution Homogenizer (Bertin, P000062-PEVO0-A). Protein concentration was measured with a Pierce BCA assay according to the manufacturer's protocol. 20-40µg of protein were prepared for western blot, depending on the protein to be examined. Samples were boiled for 5 minutes in LDS sample buffer with 2.5mM dithiothreitol (DTT), and then run on a NuPAGE 4-12% Bis-Tris protein gel in 1X MES SDS running buffer. Proteins were then transferred to a PVDF membrane in 1X transfer buffer at 4°C. Following transfer, the membrane was blocked in 5% milk blocking buffer for 1 hour, and then incubated in blocking buffer + primary antibody overnight at 4°C. The following morning, the membrane was washed in 1X PBS + 0.05% tween for 3X 15-minute washes. Proteins were then incubated with secondary antibody in blocking buffer for 1 hour, then washed three times again in 1X PBS +

0.05% tween and once in 1X PBS. For CNTNAP2, which is weakly expressed in our cultures, a Pierce enhanced chemiluminescence (ECL) kit was used to enhance protein detection (following the provided instructions). The blot was then immediately visualized with a ChemiDoc imaging system (BioRad, 17001401) and data subsequently analyzed with Image Lab 6.0.1 software (BioRad).

3.2.7 CRISPR-Cas9 genome engineering

All CRISPR gene-targeting protocols were adapted from Bruntraeger et al. (280). CRISPR RNAs (crRNAs) targeting the first few exons of *CNTNAP2* were designed with IDT's Custom Alt-R CRISPR-Cas9 design tool. The cRNA with maximal on-target efficiency and minimal off-target effects was chosen. This cRNA, denoted as *CNTNAP2_KO_exon3*, targets exon 3 of the gene and binds within the first tenth of the full-length coding sequence. To facilitate screening for positive knock-out clones, a single-stranded oligodeoxynucleotide (ssODN) repair template was designed to insert a stop codon (**TAA**) contained within a PAC1 restriction site (**AATTAATT**). NDC1.2 iPSCs with a low passage number were grown to 70-80% confluency for CRISPR nucleofection. Several hours before nucleofection, culture media was changed to E8 + rock inhibitor. Immediately before nucleofection, gRNAs were prepared by combining tracrRNA and crRNA to a final concentration of 45 μM in duplex buffer, and annealing at 95°C for 5 minutes. To form the Cas9-ribonucleoprotein (RNP) complex, 24.4 pmoles of spCas9 were combined with 45 pmoles of the crRNA:tracrRNA complex. hPSCs were dissociated with accutase, pipetted to achieve a single cell suspension, and then passed through a 50 μm cell strainer. hPSCs were then counted and 200,000 cells per nucleofection centrifuged at 400 r.c.f. for 4 minutes. The resulting supernatant was removed and the cells re-suspended in P3 solution, Cas9-RNP, and 100 μM repair ssODN. hPSCs were nucleofected using an Amaxa 4D nucleofector (Lonza, AAF-1002B/X), and immediately transferred to a geltrex plate containing E8 + rock inhibitor. Thereafter, culture media was changed every 24 hours to fresh E8 without rock inhibitor.

Once hPSCs reached 70-80% confluency, they were dissociated with diluted EDTA and split one-third for DNA collection, two-thirds for further passaging. Population-level DNA was collected to determine overall gene targeting efficiency through 1) a T7 endonuclease I assay and 2) a PacI restriction digest. For both assays, DNA was isolated with a DNeasy blood and tissue kit (Qiagen, 69506) and used as a PCR template to amplify a ~500bp region surrounding the Cas9 cut site (KAPA HiFi HotStart ReadyMix, KR0370). For the T7 assay, the PCR products were hybridized in 10X NEBuffer 2 as follows: 95°C: 10 minutes, 95-85°C: ramp -2°C per second, and 85-25°C: ramp -0.3°C per second. The resulting duplex DNA was then incubated with T7 endonuclease I for 1 hour at 37°C before running on a gel. For the PacI restriction digest, the PCR products were incubated with PacI enzyme for 1 hour at 37°C and visualized on a gel. A T7 gel with multiple bands indicates the presence of heteroduplex DNA, and likely CRISPR editing events. A PacI gel with multiple bands suggests homozygous or heterozygous insertion of the TAA stop codon.

For the hPSCs that were further passaged, half were frozen down and half split again for single cell plating. Once hPSCs reached 70-80% confluency, they were dissociated with accutase, pipetted to achieve a single cell suspension, and passed through a 50 μm cell strainer. Cells were then counted and serially diluted to a final concentration of 10,000 cells/mL. Approximately 1000-1500 clones were plates on a 10cm geltrex-coated dish containing cloning media. 48 and 72 hours post-plating, culture media was replaced with fresh cloning media, and thereafter every 24 hours with E8. When 1-2mm colonies appeared, hPSCs were picked for single cell screening. Using a tissue culture microscope, 60 colonies were transferred to new geltrex-coated plates containing E8 + rock inhibitor. Once at 70% confluency, colonies were dissociated with diluted EDTA and half of each colony frozen down, half harvested for genomic DNA. Clone-level DNA was isolated with QuickExtract DNA extraction solution, and PCR amplified for the same 500bp region surrounding the Cas9 cut site.

Putative knock-out clones were identified by 1) PacI digestion, 2) sanger sequencing, and 3) MiSeq next generation sequencing. Sanger sequencing was performed by the Department of Biochemistry (University of Cambridge) sequencing service. MiSeq was performed by the Sanger Institute. To confirm knock-out at the RNA and protein level, lines from two putative knock-out clones were differentiated to neurons (*CNTNAP2* is not expressed in hPSCs). At day 50, RNA and protein were harvested for analysis with qPCR and western blot.

3.2.8 Single cell RNA sequencing (scRNA-Seq)

Day 50 cultures were washed once with 1X PBS then incubated with papain dissociation mix for 10 minutes at 37°C and 800 RPM. The cells were mixed by gentle pipetting, then returned to the heat block for another 10 minutes. Following the second incubation, the cells were centrifuged at 300 r.c.f. for 4 minutes and resuspended in inhibitor mix, ovomucoid, and EBSS. The cell suspension was then centrifuged again for another 4 minutes at 300 r.c.f., and resuspended in ice-cold 1X PBS + 0.5% BSA . The cells were then filtered through a 50µm cell strainer and counted. 2 million cells per sample were aliquoted into a low-bind Eppendorf tube, spun cold at 350 r.c.f. for 6 minutes. The resulting pellets were resuspended in 100 µL Cell Staining Buffer. After resuspension, 5 µL FcR Block was added to the cells, and the solution incubated on ice for 5 minutes. 1µg of cell hashing antibody was added to each sample (different antibodies per sample). The cells were then incubated at 4°C for 30 minutes.

After hashtag antibody incubation, the samples were washed three times in 1.2 mL Cell Staining Buffer and spun at 350 r.c.f. and 4°C for 6 minutes. The pellet was then resuspended in 400uL EBSS + 0.04% BSA. Each sample was then counted and pooled together at a concentration of 500,000 cells per sample. The scRNA-Seq library was then prepared according to the Chromium Next GEM Single Cell 3′ Kit v2 protocol (10X Genomics). After cDNA amplification, HTO-derived cDNA (~200 bp) was

separated from mRNA-derived cDNA (~400 bp) using size-selection beads called SPRI-select. After size selection, HTO-derived cDNA were prepped according to the cell hashing protocol from Stoeckius et al. (281), while the mRNA-derived cDNA continued with the 10X genomics kit. Shortly before sequencing the HTO-derived library and mRNA-derived library were combined to a final concentration of 4 nmol/L. For sequencing, a 100 cycle run was performed on an Illumina NovaSeq to obtain approximately 800 million reads across the pooled sample. The sequencing data was aligned/processed with the Cell Ranger pipeline (10X Genomics) and analyzed with the R package Seurat version 3.1 (282). Overall, 18,304 cells were sequenced, with an average of 51,780 reads per cell mapping to a median of 1,818 genes per cell.

3.2.9 CNTNAP2 knock-out culture measurements

Knock-out hPSCs were simultaneously differentiated alongside the un-edited parental hPSCs (NDC1.2) using identical reagents and handling. Only if the two cultures showed similar transcriptional profiles on the nCounter SPRINT Profiler, were they used in downstream experiments. On day 35 cultures were sparsely transfected with pSynapsin1-mNeonGreen or pCamkii-mKate2 (i.e. green or red). Sparse labelling was achieved as follows: cells were dissociated and single celled with accutase, filtered through a 50µm cell strainer, washed with NMM. 500,000 cells were counted and spun down, and the resulting pellet re-suspended in P3 solution. Cells were then transfected with 3µg of the designated plasmid using an Amaxa 4D nucleofector (Lonza, AAF-1002B/X). Nucleofected cells were immediately mixed 1:1 with either 500,000 non-transfected cells of the same genotype, or with 500,000 cells of the opposite genotype that was labelled with a different fluorophore. The cultures were then plated on a geltrex-coated 96-well imaging plate at 150,000 cells/cm^2. 24 hours post-nucleofection and every 48 hours thereafter culture media was replaced with fresh NMM. Nucleofection was chosen over low-titre lentivirus transduction, as it allows for “birth dating” of the cells being examined. With transfection, fluorescence

gets diluted out of any dividing NPCs present at day 35, and only remains in early post-mitotic neurons present at the time of labelling (i.e. deep-layer neurons of comparable age).

Dendritic spine analyses

Knock-out and wild type cultures were fixed at day 50 and day 75 with 4% PFA + 4% sucrose as described in **Methods 2.2.8**. Day 50 was chosen to capture early spine formation, while day 75 was the latest time point with useable fluorophore signals to capture mature spines and neurons. Fixed cultures were imaged on an Olympus FluoVIEW FV1000 confocal microscope with a 60X objective and 1.2X zoom. Neurons to acquire were selected randomly, but must have had a strong fluorescent signal and been easily distinguishable from surrounding cells. Wherever neurons could not be fit within the imaging frame, as much of the cell as possible was acquired. Z-stacks of the selected neurons were taken to completely image all dendrites and spines. Cultures were imaged at 8 pixels per second and processed with a Kalman line filter. Images were then analyzed with the semi-automated NeuronStudio software to measure dendritic spine density.

Neurite analysis

Between day 40 – 70 the cultures were imaged live every ~5 days on a PerkinElmer Opera Phenix microscope (fitted with a 37°C/5% CO_2 imaging chamber). The Opera Phenix was automated to image the centre of each well, acquiring large quantities of neurons in a non-biased fashion. Using NeuronStudio, the following parameters were measured from the images: 1) dendritic branching, 2) total neuron length, and 3) average neurite length per neuron.

3.2.10 Immunofluorescence staining

Culture media was aspirated and cells washed with 1X PBS. Cells were fixed for 15 minutes at room temperature with 4% PFA + 4% sucrose solution. Fixed cells were washed three times with 1X TBS for 5 minutes, then washed another three times in 1X TBS with 0.3% Triton X100. Cells were blocked in TBS + Triton X100 with 5% normal goat serum for 1 hour. Primary antibodies were diluted in blocking buffer and incubated with cells overnight at 4°C. The following day, cells were washed three times with 1X TBS, three times with TBS + Triton X100, and incubated with Alexa Fluor secondary antibodies for 2 hours (diluted 1:500 in blocking buffer). Finally, cells were washed six times with 1X TBS. 1µg/mL DAPI solution was added to the third wash. Stained cultures were imaged on either a PerkinElmer Opera Phenix or Olympus FluoVIEW FV1000 confocal microscope.

3.2.11 Luciferase enhancer assays

Based on Prescott et al. (2015), a luciferase assay was designed test the enhancer potential of the *CNTNAP2* HARs. Six of the HARs; CNS97, CNS116, 3xHAR.395, CNS590, CNS884, and CNS954, were cloned into a pGL3 vector containing a firefly luciferase gene under the control of a SV40 promoter (Promega). If a HAR had enhancer or repressor potential, the presence of the HAR would be hypothesized to increase/decrease luciferase signal relative to a control plasmid with no HAR present. Genomic DNA isolated from NDC1.2 hPSCs was used to PCR-amplify the six HARs with a Phusion high-fidelity DNA polymerase. The PCR products were then run on a gel and purified using a Wizard SV Gel and PCR Clean-Up kit (Promega, A9285). The pGL3 vector was digested with XhoI before each of the HARs were inserted into the vector using in-fusion cloning. Correct assembly of the vectors was confirmed by sanger sequencing each plasmid with the RV3 and GL2 sequencing primers.

SH-SY5Y cells were partially thawed in a 37°C water bath, washed in DMEM + glutamax, and centrifuged at 400 r.c.f. for 4 minutes. The cell pellet was re-suspended in STO media and grown in a Nunc EasYFlask with a filter lid. Once cells were confluent, they were washed with 1X PBS and dissociated in TrypLE for one minute. Once SH-SY5Y cells detached, they were collected in PBS, spun down, and re-suspended in fresh STO media. Cells were passaged at least once before vector transfection. Once at confluency, FuGENE HD was used to transfected SH-SY5Y cells with 1000ng of each pGL3 vector and 50ng of renilla luciferase normalization vector (pRL). FuGENE was diluted in Opti-MEM media and mixed with plasmids in a 2.5 FuGENE : 1 DNA ratio according to the manufacturer's instructions. STO media was replaced the following morning, and 48 hours post-transfection SH-SY5Y cells were prepared for the luciferase assay according to the provided protocol. Firefly and renilla luciferase signals were measured for each HAR using a Hidex sense microplate reader on the luminometer setting.

3.2.12 Neutrality tests

The Tajima's D test and the integrated haplotype score (iHS) test were performed on a server at the High Performance Computing Centre, University of Tartu (168). All genome data was pre-processed by others on the Tartu server, as outlined in Pagani et al. (283). The genomes analyzed were part of the Estonian Biocentre human Genome Diversity Panel (283). This panel contains 483 high-coverage human genomes from 148 populations worldwide (see **Chapter 7** for full details). PLINK (v1.90) and VCF Tools (v0.1.14) were then used to run Tajima's D in 2000 base pair windows for each population. The iHS test was by Pagani et al. (283) – please refer to the article for a detailed description of their methods.

The Combined Annotation Dependent Depletion (CADD) scores (v1.6) were taken from the publicly available dataset at: https://cadd.gs.washington.edu/ (284, 285). This model will be described in detail in **Chapter 7**. In brief, however, it assigns

a ‘deleterious score’ to every possible nucleotide substitution in the human reference genome (build GRCh38/hg38).

3.2.13 Statistical analyses

All statistics analyses were performed with either ‘R Studio’ (v1.2.5033) (286) or ‘GraphPad Prism’ (v8.4.2). The following packages were utilized in R for data analysis/presentation:

i) tidyverse (v1.3.0) (https://cran.r-project.org/web/packages/tidyverse/index.html)
ii) ggpubr (v0.4.0) (https://cran.r-project.org/web/packages/ggpubr/index.html)
iii) rstatix (v0.6.0) (https://cran.r-project.org/web/packages/rstatix/index.html)
iv) dplyr (v1.0.0) (https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html)
v) ggplot2 (v3.3.1) (https://cran.r-project.org/web/packages/ggplot2/index.html)
vi) pheatmap (v1.0.12) (https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf)
vii) Seurat (v3.1.5) (282)

Chapter 4

CNTNAP2 expression in human and macaque *in vitro* forebrain neurons

4.1 Introduction

As discussed in **Chapter 2**, the *CNTNAP2* gene is highly expressed in the human fetal cortex (197, 287). Studies of rodent fetal brain, however, show a strikingly different pattern with almost no cortical expression (197, 245). These differences reinforce that while rodent studies on cortical development can be extremely useful, caution must be taken when extrapolating their findings to humans. This is especially important for 'brain evolution genes', as many are hypothesized to have altered expression between species (25). Unfortunately, the vast majority of existing *Cntnap2* studies have been conducted on mice – meaning little is known about the

temporospatial changes to *CNTNAP2* expression during human corticogenesis. The few existing 'human' expression studies are limited in number, provide only a snapshot of one developmental timepoint, and/or are not optimized for *CNTNAP2* (as with transcriptome-wide approaches like RNA-Seq) (3, 197, 245, 260, 287).

Understanding when, where, and how *CNTNAP2* is expressed in the human cortex will provide critical clues about its function and its role in disease. This is especially important considering the evidence that *CNTNAP2* dosage contributes to disease severity. In particular, individuals with homozygous mutations have more serious phenotypes than those with heterozygous mutations (e.g. cortical dysplasia focal epilepsy (CDFE) or Pitt-Hopkins syndrome (PTHS)) (189). The precise pathogenic mechanisms of how changes to *CNTNAP2* dosage lead to such phenotypes are still largely unclear. Answering these questions will be essential to finding effective treatments/preventions for these and other patients. Provided the potential role of *CNTNAP2* in human evolution - and the importance of differential gene expression for human-primate differences (25) - it may also shed light on what makes our species 'human'.

Given the practical and ethical challenges of using human brain tissue, expression studies with primary samples are not always possible (**Chapter 1**). While animal models have historically been used as an alternative, they may not accurately recapitulate human disease mechanisms. For instance, many drugs that are effective in mice have subsequently failed in human clinical trials (288). Their use in investigating human evolution is also not ideal, as they lack human-specific biology. Pluripotent stem cell (PSC)-derived *in vitro* models have been used as an effective substitute for over a decade (113, 173, 181). This approach allows experiments to be carried out across corticogenesis, in a human biological context, and over larger scales with relative ease. Finally, transcriptomic comparisons of PSC-derived tissue with their primary orthologues have revealed they are faithful models (3, 4, 113, 173, 181, 289).

In this chapter, I report data on the analysis of *CNTNAP2* expression over the course of human *in vitro* cortical development (using our PSC-derived model system). I also compare *CNTNAP2* expression between human and non-human primate *in vitro* cortical cultures. Finally, I introduce a human PSC-derived *CNTNAP2* knockout line, and the methods used to generate and validate it.

4.1.1 Quality control of cortical inductions from pluripotent stem cells

Before investigating *CNTNAP2* expression in our PSC-derived cortical cultures, I first sought to 'quality control' the inductions by confirming they were of correct regional identity. This would serve two purposes: 1) ensuring *CNTNAP2* expression was indeed measured in human cortical neurons, and 2) allowing culture maturity and composition to be correlated back to *CNTNAP2* expression. To do so, RNA from every induction was run on a **Nanostring nCounter SPRINT Profiler** (hereafter referred to as the 'Nanostring') (290). With this platform, mRNA from user-defined genes of interest are tagged by two probes (see **Figure 4.01**). The first, is a **fluorescently barcoded 'reporter probe'** (one barcode per gene). The second, is a **biotinylated 'capture probe'**. The capture probe attracts the target mRNA to a **streptavidin-coated** imaging plate. The plate is then scanned by an automated fluorescent microscope that counts the number of barcodes present for each gene. The barcodes are thus a proxy for the number of mRNA molecules in the culture. The Nanostring offers several advantages over traditional gene expression assays such as qRT-PCR. Not only is it possible to test hundreds of genes simultaneously, but it bypasses the mRNA to cDNA conversion step that can introduce bias into the measurements (290). With the Nanostring, mRNA molecules are counted directly instead.

The ~150 genes detected by the Nanostring (see the **Appendix** for a list of the probes and their sequences)encompass **1) positive controls:** genes expressed by telencephalic cell types (e.g. progenitors, cortical neurons, and cortical glia), **2) negative controls:** genes expressed by non-telencephalic cells (e.g. stem cells, diencephalon, midbrain, or hindbrain), and **3) genes for normalization:** house-keeping genes to control for differences in mRNA input. This platform was designed and tested to enable high-throughput comparison of culture maturity, relative cellular composition, and regional identity (187). Inductions were considered successful if they showed robust expression of cortex/telencephalic genes, and weakly expressed markers of the negative controls. Empirical thresholds of expression were set based

on ~160 previous inductions in the Livesey lab (see Strano et al. for full details (187)).

For my purposes, RNA was taken from four key stages of the induction process: **1) stem cell**, **2) neural progenitor cell** (~D30), **3) deep layer formation** (~D30), and **4) upper layer formation** (~D80). These timepoints were chosen because they overlap with major milestones during corticogenesis. To account for variability between inductions and cell lines, one induction from three human PSC lines were used: **1) H9** (embryonic PSC), **2) NDC1.2** (induced PSC), and **3) SFC840** (induced PSC).

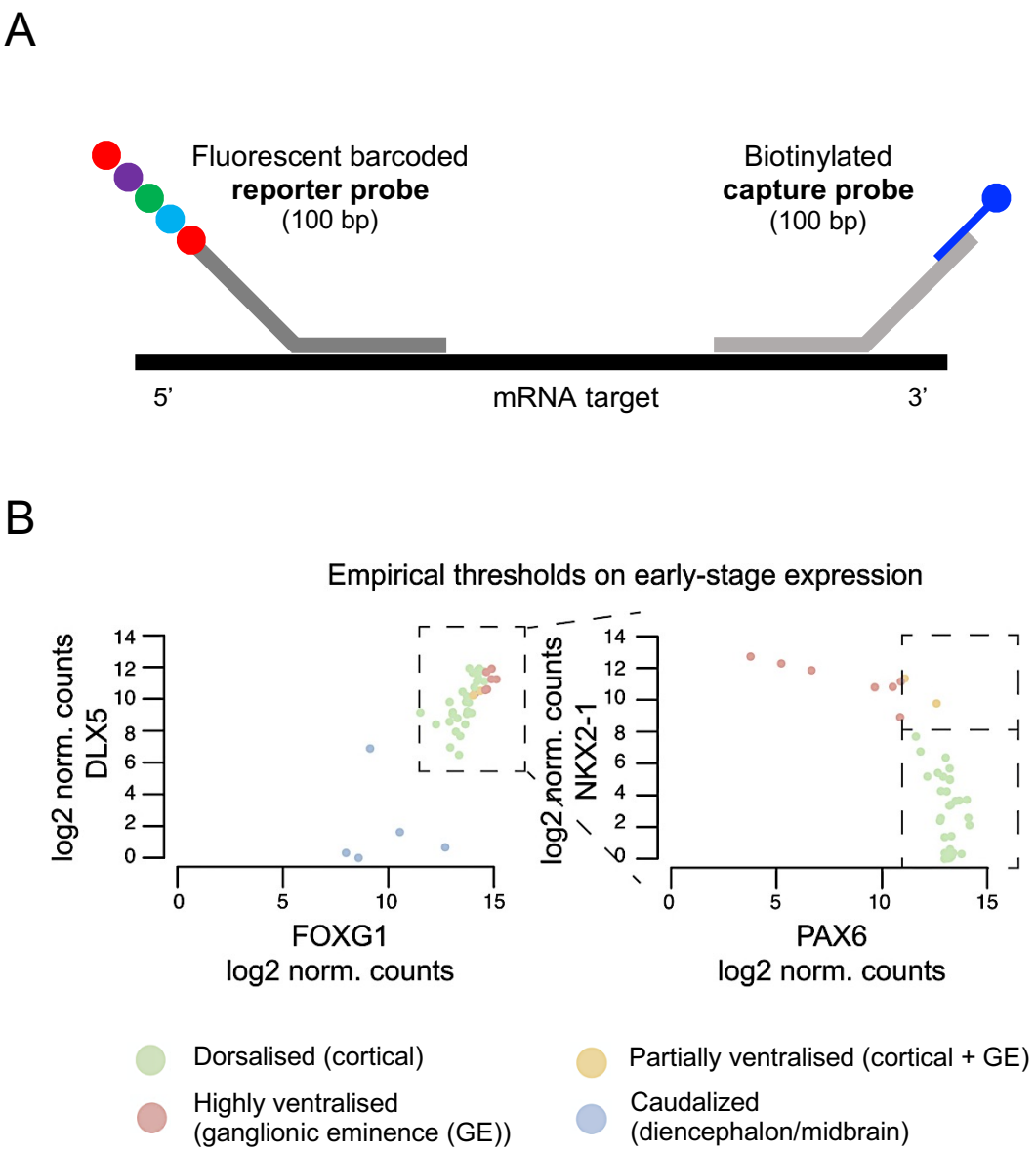


Figure 4.01. The Nanostring nCounter SPRINT Profiler

(A) In the Nanostring assay, target mRNA is tagged by two probes: 1) a reporter probe (containing a gene-specific fluorescent barcode) and 2) a capture probe (which is biotinylated and attracts the mRNA to a streptavidin-coated imaging plate). A fluorescent microscope then counts the number of barcodes present for each gene. (B) Example classifications of cortical differentiations based on Nanostring data. The specific thresholds on the expression of positive control genes (e.g. those expressed by cortical neurons) and negative control genes (e.g. those expressed by diencephalon or midbrain) are set empirically based on 160 previous inductions. For full details please see Strano et al. (2020) (source of figure B).

4.2 *CNTNAP2* expression during *in vitro* human corticogenesis

4.2.1 Results of the Nanostring – expression of neuronal genes across culture development

A clear reduction in stem cell and pluripotency markers is observed from the PSC stage onwards (**Figure 4.02b**). Similarly, genes associated with neural progenitor cells are strongly up-regulated by D30 (**Figure 4.02c**). These genes continue to be transcribed into the upper layer formation stage (~D80), indicating continued progenitor proliferation. For example, the radial glia-associated gene, ***PAX6*** (181), is significantly up-regulated during culture development [one-way repeated measures ANOVA, $F(3, 6) = 241.98$, $p = 1.2e\text{-}6$]. Pooling together the three inductions, post-hoc paired *t*-tests confirmed that the increase from PSC to D30, PSC to D50, and PSC to D80 were significant after Bonferroni correction ($p = 0.0005$, $p = 0.009$, and $p = 0.01$ respectively). There was no significant difference between D30, D50, and D80.

Excitatory and inhibitory cortical markers can also be seen from the neural progenitor cell (NPC) stage onwards (**Figure 4.02d-e**). Deep layer neurons are present from ~D30 (as shown by the presence of *TBR1*, a deep-layer specific gene) while expression of upper layer-specific markers (e.g. *SATB2*) is still low at ~D80. For *TBR1*, ANOVA found a significant effect of culture age [$F(3, 6) = 45.45$, $p = 0.0002$]. This was revealed to be a significant increase from PSC to D30 ($p = 0.004$) and from PSC to D80 ($p = 0.008$). No difference was detected for PSC to D50, or any of the remaining pairwise combinations. For *SATB2*, conversely, ANOVA found no significant change in expression with time [$F(3, 6) = 3.29$, $p = 0.1$]. This suggests upper layer neurons are not yet present in large quantities at D80.

Later formed interneuron types such as parvalbumin (PV^+) and somatostatin (SST^+) are also not observed by ~D80. This is despite clear expression of interneuron lineage markers such as *DLX1* and *GAD1*. Checking *DLX1* expression across culture age revealed a significant change [$F(3, 6) = 14.18$, $p = 0.004$]). However, this significance did not survive correction for multiple testing in post-hoc analyses ($p > 0.05$ for all). As this trend was noted for other interneuron-associated genes, and interneuron genes should not be expressed in PSCs, the number of inhibitory cells in the cultures is likely to be low.

Finally, non-neuronal (but cortical) cell types, such as oligodendrocytes and astrocytes are observed between ~D50 - D80 (**Figure 4.02f**). They are likely to be present in only minor quantities, however, as ANOVA did not find the expression of *OLIG1* (an example oligodendrocyte-associated gene (291)) to be significantly different with time [$F(3, 6) = 3.59$, $p = 0.09$].

Taken together, these findings suggest the three PSC cell lines have been successfully differentiated to a dorsal telencephalic/cortical fate. However, they are still relatively immature, as evident from the lack of upper layer gene expression. Non-telencephalic cell types (e.g. diencephalon, midbrain, and hindbrain) appear to be absent, and non-dorsal telencephalon (i.e. ganglionic eminence or interneurons) are low in number. A more detailed analysis of culture composition will be performed with single cell RNA sequencing (scRNA-Seq) in **Chapter 5**.

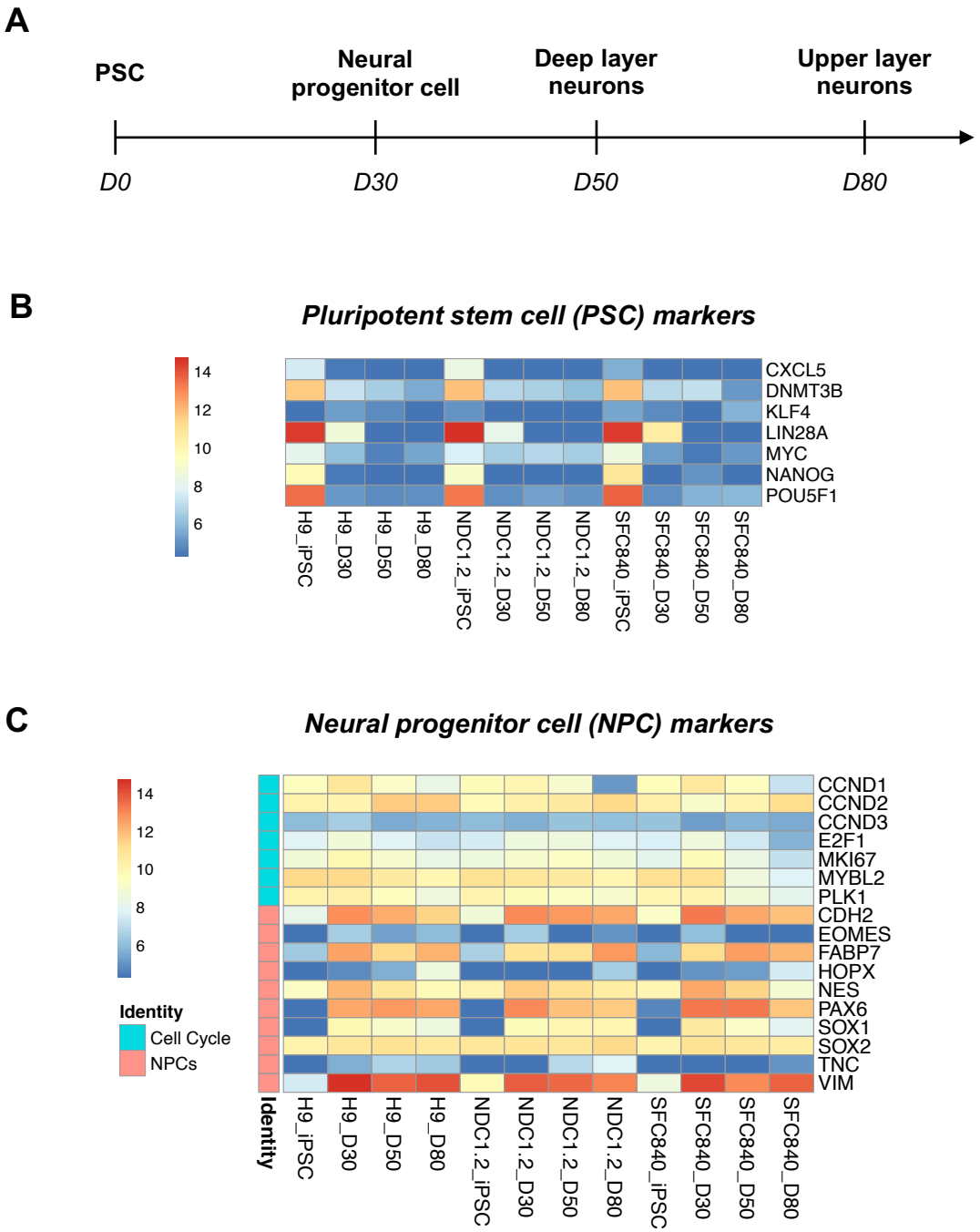


Figure 4.02. Transcriptional profiles of human *in vitro* cortical inductions (figure 1 of 3)

Heatmaps show log2 normalized RNA counts of each gene measured by the Nanostring. One induction per cell line was used with each time-point representing 50ng total RNA. (A) Approximate timeline of the Shi et al. differentiation protocol. (B) Stem cell genes are highly expressed in PSCs, and decrease post-differentiation. (C) Cell cycle genes are detected at all time-points, indicating cell division is occurring. Neural progenitor-associated markers are detected in all stages except for stem cells, further supporting that progenitor proliferation is ongoing throughout the 80 sampled days.

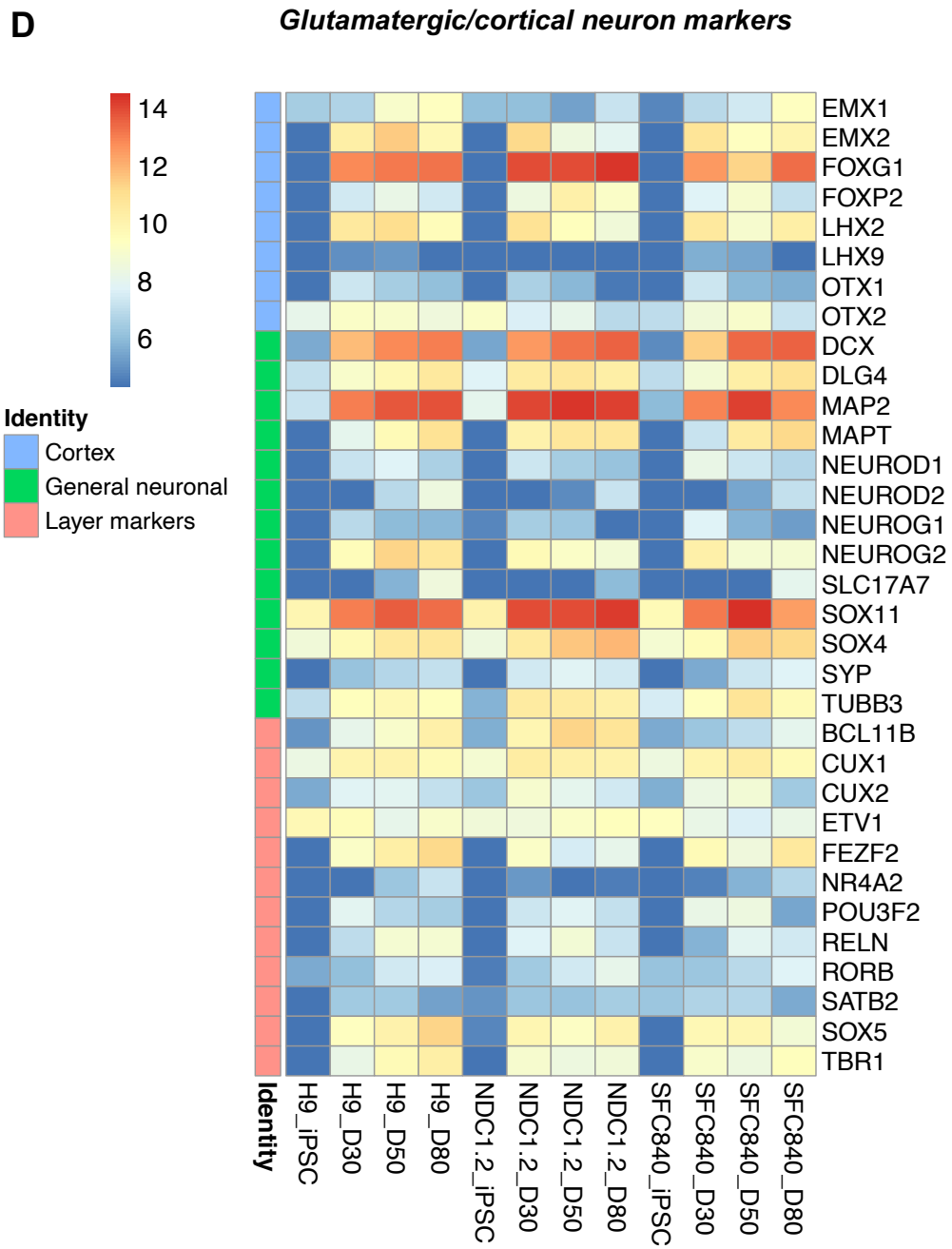


Figure 4.02. Transcriptional profiles of *in vitro* cortical inductions (2 of 3)

Heatmaps show log2 normalized RNA counts of each gene measured by the Nanostring. One induction per cell line was used with each time-point representing 50ng total RNA. (D) Genes are grouped into three broad categories ("identities"): cortex-specific genes, general neuronal genes, and cortical layer markers. Almost all genes show a trend of increased expression with culture development, and are notably absent in stem cells. Deep layer markers are detected from ~D30, while upper layer markers are not yet robustly expressed by ~D80.

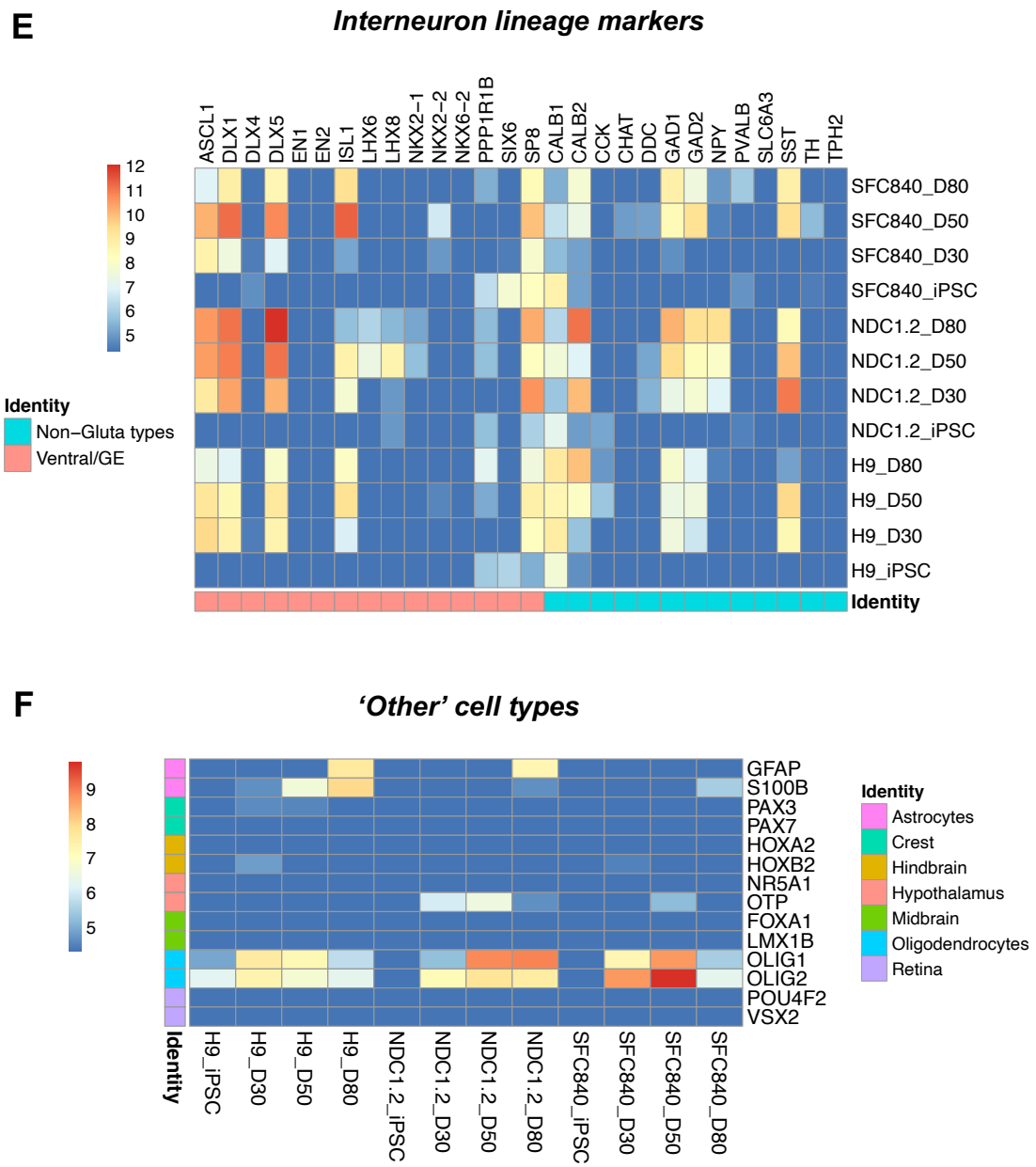


Figure 4.02. Transcriptional profiles of *in vitro* cortical inductions (3 of 3)

Heatmaps show log2 normalized RNA counts of each gene measured by the Nanostring. One induction per cell line was used with each time-point representing 50ng total RNA. (E) Genes are grouped into two categories ("identities"): genes expressed in ganglionic eminence (GE; where cortical interneurons are formed), and GABAergic lineage markers (non-glutamatergic). In general, both categories of genes are weakly expressed throughout culture development. However, expression of certain markers (e.g. *DLX5*) is detected, indicating low proportions of inhibitory neurons are present in the culture. (F) Oligodendrocytes are observed from D30, while astrocytes are present at later time-points (e.g. D80). Cell types produced by unsuccessful inductions (i.e. non-cortical) are largely absent from the culture makeup.

4.2.2 CNTNAP2 mRNA expression time-course analysis

Having examined the transcriptional profiles of three cortical differentiations, I then set out to measure *CNTNAP2* mRNA and protein expression. This was achieved using standard qRT-PCR and western blot assays (**Chapter 3**). As before, RNA was taken at four developmental stages: 1) PSC, 2) NPC (D30), 3) deep layer neuron (D50), and upper layer neuron (D80). This experiment was also replicated with three different cell lines, derived from different donors/genomes, including one ES cell line (one induction per line): 1) AD2.1, 2) H9, and 3) NDC1.2.

qRT-PCR showed a clear increase in *CNTNAP2* mRNA expression over time, starting from no detectable expression in stem cells to the strongest expression in D80 cultures (**Figure 4.03**). One-way repeated measures ANOVA revealed *CNTNAP2* levels significantly increased with culture age [$F(3, 24) = 167.851$, $p = 3.1e\text{-}16$]. Post-hoc analyses subsequently confirmed the PSC to D30 transition ($p = 1.00e\text{-}3$) and the D50 to D80 transition ($p = 4.11e\text{-}5$) were significantly increased (pairwise *t*-tests with Bonferroni correction). Detectable expression at D30 in culture, before neurons are generated, supports previous findings suggesting *CNTNAP2* is expressed in cortical progenitor cells, as well as in neurons. Higher expression at later stages is consistent with higher expression in post-mitotic neurons.

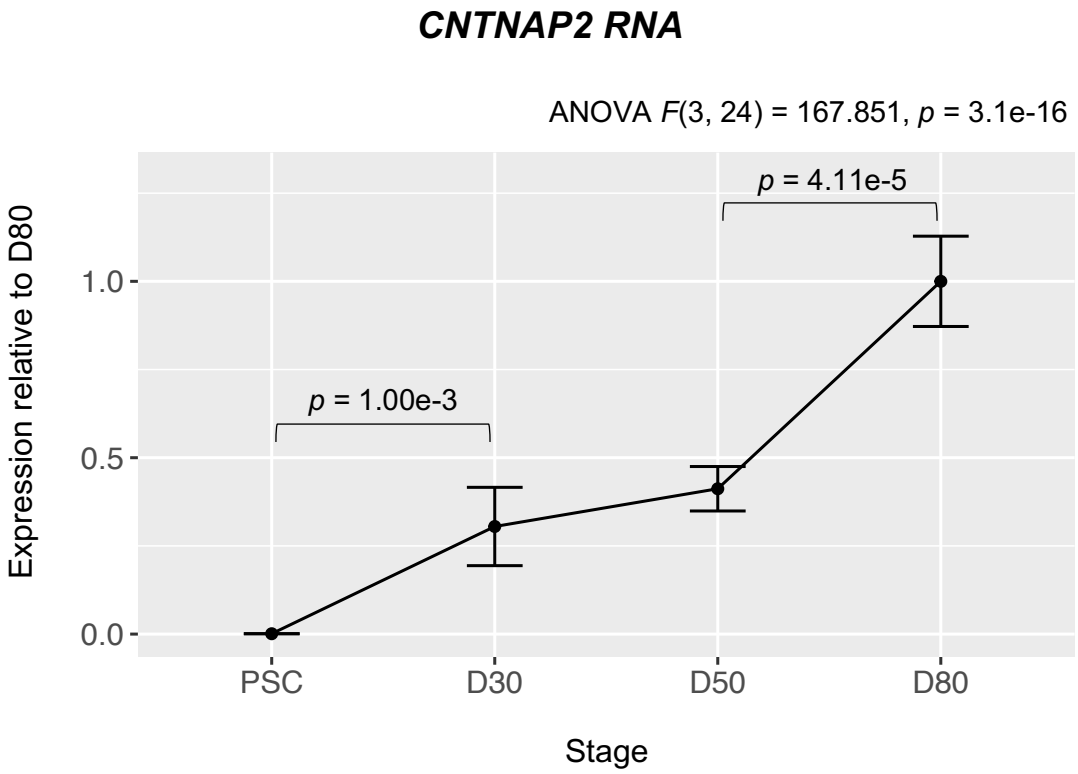


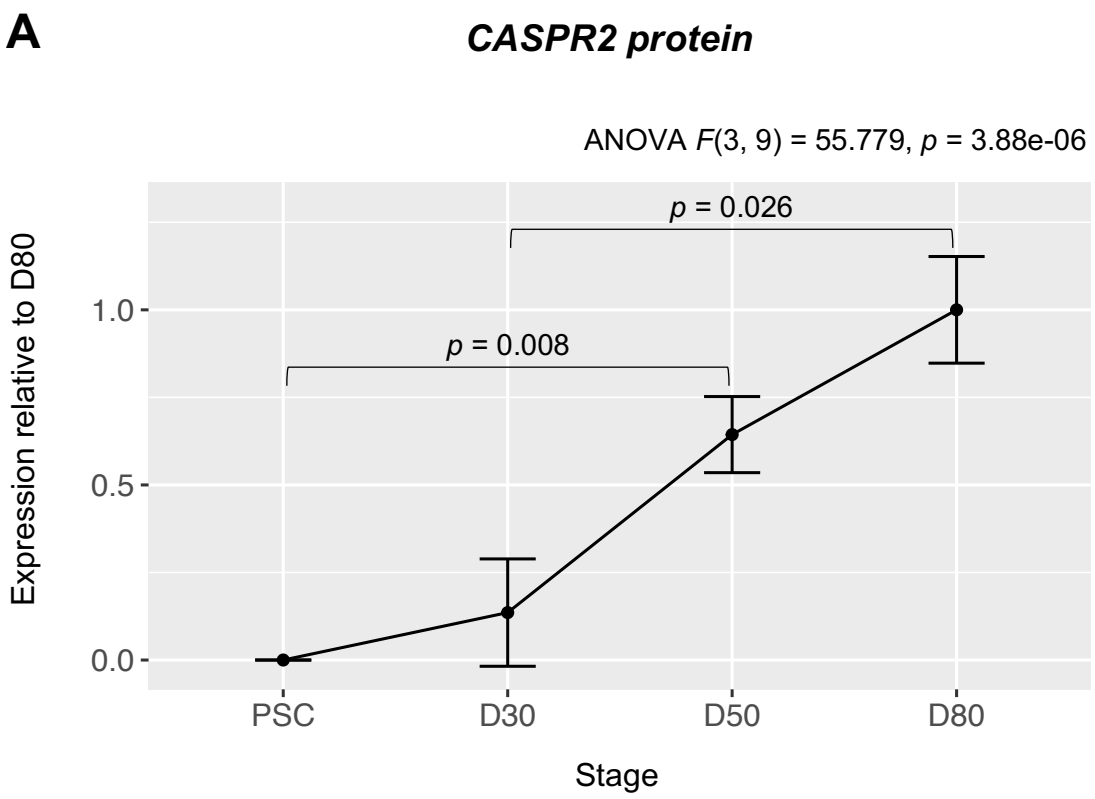
Figure 4.03. mRNA expression of *CNTNAP2* during human in vitro cortical development

qRT-PCR results from three cell lines at four stages across culture development (one induction per line, three replicates per timepoint). Significant increases in *CNTNAP2* expression were noted between stem cell to D30, and D50 to D80 (repeated measures ANOVA; post-hoc paired t-tests adjusted with Bonferroni correction). 200 ng of total RNA was used for each sample. Data was normalized to the geometric mean of two housekeepers, *RAB7A* and *C1ORF43*. Error bars represent 95% confidence intervals; adjusted p-values shown.

4.2.3 CASPR2 protein time-courses

With the finding that *CNTNAP2* mRNA expression increases with culture development, I next wanted to confirm whether the same trends could be seen at the protein level (as a reminder, the *CNTNAP2* gene produces CASPR2 protein). I therefore extracted protein at each of the four timepoints previously mentioned: 1) PSC, 2) NPC (D30), 3) deep layer neuron (D50), and upper layer neuron (D80). Two cell lines were used (one induction per line): 1) AD2.1 and 2) SFC840. All western blots were performed with a KO-validated CASPR2 antibody that I had pre-optimized for use in our cultures (NeuroMab 73-075).

Like with qRT-PCR, the western blot results showed a clear increase in CASPR2/*CNTNAP2* protein expression over time. There was no expression in stem cells, variable (but present) expression in D30 NPCs, and more robust expression by D50 and D80 (**Figure 4.04**). One-way repeated measures ANOVA confirmed CASPR2 expression significantly increased with time [$F(3, 9) = 55.779$, $p = 3.88e\text{-}06$]. Pairwise *t*-tests revealed PSC to D50 ($p = 0.008$) and D30 to D80 ($p = 0.026$) were significantly increased after Bonferroni correction.



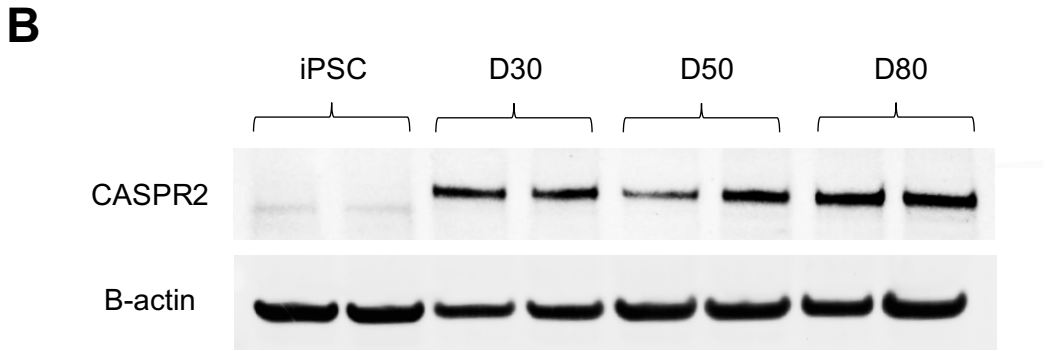


Figure 4.04. Protein expression of CASPR2 during human in vitro cortical development

(A) Western blot results from two cell lines (one induction per line, two replicates per timepoint). As with qPCR, CASPR2 expression increased with culture maturity. However, the increases between D30-D50 and D50-D80 were not statistically significant. All measurements were normalized to B-actin. (B) Representative western blot from an AD2.1 induction. 40 µg of protein was loaded for each sample. The CASPR2 protein has a molecular weight of 148 kDa, and was detected as a single band on the protein gel at ~150 kDa, confirming its identity.

4.2.4 Cellular localization of the CASPR2 protein

In addition to studying the relative levels of *CNTNAP2*/CASPR2 expression, I also wanted to examine the gene's cellular location. In mice, CASPR2 is expressed as **puncta** along axons, dendrites, and soma (34, 202, 203). Importantly, these puncta co-localize with excitatory and inhibitory synapses (203). This localization provides further suggestion that *CNTNAP2* is important for synaptic function. Whether this pattern is also observed in humans, however, is unclear.

To address this, I used immunofluorescence with antibodies against CASPR2, **TAU** (axons), **MAP2** (dendrites), and **DAPI** (a nuclei stain) (292). To accurately determine whether CASPR2 was localized to synapses, I also stained for two synaptic proteins: **PSD-95** (a **post-synaptic** scaffolding protein (293)) and **SYNAPTOPHYSIN** (a **pre-synaptic** vesicle protein (294)).

As shown in **Figure 4.05,** human NDC1.2 neurons (fixed at D70) show discrete CASPR2 puncta along axons, dendrites, and soma (i.e. similar to what is seen in mice). Puncta co-localized with both pre-synaptic and post-synaptic markers, confirming their location at the synapse. Although not exhaustive testing, these results indicate the cellular pattern of CASPR2 expression is likely to be similar between humans and other species. It also provides further evidence that *CNTNAP2* is involved in synaptic transmission in humans, and that *CNTNAP2* is expressed by neurons in our culture system.

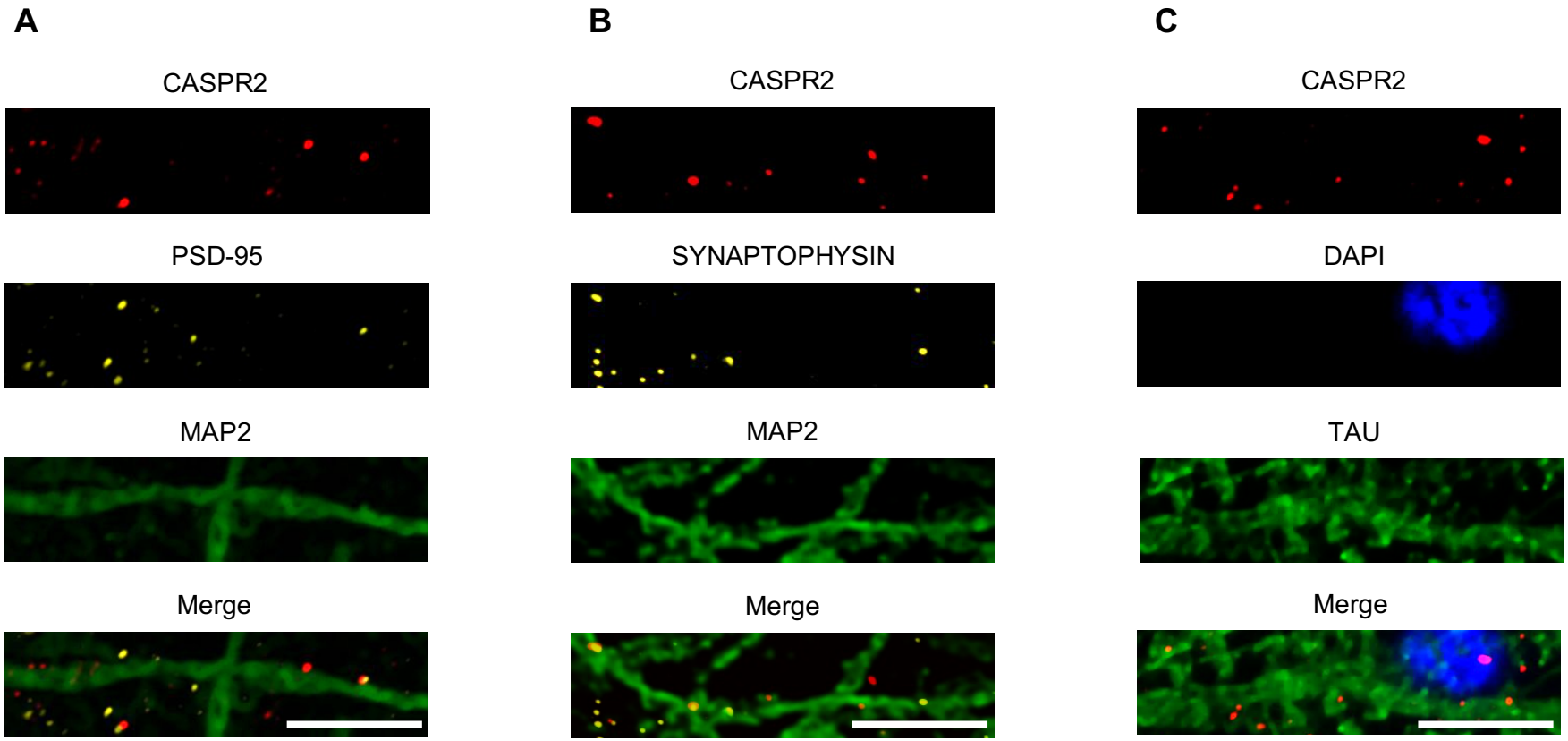


Figure 4.05. Cellular expression of CASPR2 in human PSC-derived forebrain neurons (D70, NDC1.2 cell line)

(A-B) CASPR2 is expressed as puncta along dendrites (which are stained by MAP2). Puncta colocalize with (A) post-synaptic protein PSD-95, and (B) pre-synaptic protein SYNAPTOPHYSIN. (C) Puncta are also noted along axons (TAU) and soma (DAPI). Scale bar represents 5 μm.

4.3 Comparison of *CNTNAP2* expression in human and macaque PSC-derived cultures

As reviewed in **Chapter 2**, differences in the expression of *CNTNAP2* have been previously documented between human and non-human primate cortex. Nowick and colleagues (37) found a 1.3-fold increase in *CNTNAP2* expression in adult human pre-frontal cortex compared to chimpanzees. This finding was then corroborated by Muntane et al. (39), who also showed increased expression on the human-chimpanzee lineage (in comparison to Gibbons, Old and New World Monkeys, Lemurs, and Lorises). That said, other studies have not detected such differences. In particular, a scRNA-Seq study by Pollen et al. (3) found no difference in *CNTNAP2* expression between human and macaque primary fetal cortex (post-conception weeks 9-22). In PSC-derived cerebral organoids, the authors also found *CNTNAP2* was more highly expressed in chimpanzee neurons than in human neurons (i.e. the opposite direction).

There are a number of limitations to these existing studies. Firstly, no study has been conducted at the protein level - only at the mRNA level. Most of the studies also have extremely limited sample sizes and are done in bulk tissue. As mentioned, there are also reports of conflicting findings (3). Some of these shortcomings are easier to solve than others (e.g. accessing primary tissue from non-human primates is difficult). That being said, there is still great utility in performing additional studies by independent experimenters. These would offer further support for the validity of the existing studies, while hopefully providing new insights in the process.

My next step was therefore to perform differential expression analyses of human and non-human primate cortical tissue. Ideally this would involve several species (e.g. chimpanzee, gorilla, and macaque) and be conducted over several developmental timepoints (e.g. fetal, infant, juvenile, and adult). To address this, I was kindly donated PSC-derived cortical inductions from two cell lines of crab-eating

macaque (*Macaca fascicularis* – lines MF1 and MF12)[5]. These lines were differentiated using the same Shi et al. (181) protocol as our human PSC lines (4). Wherever possible they were simultaneously grown with matched human samples.

As with my *CNTNAP2* time-course experiments, I first used the Nanostring to compare the transcriptional profiles of these macaque and human cultures. This would ensure *CNTNAP2* expression would be analyzed in inductions of comparable maturity and composition. To account for species differences in gene sequence, only genes measured by Nanostring probes with >95% similarity in macaque were included (95/100 nucleotides). Three inductions from each species were used for comparison – **human:** 1) H9 (embryonic PSC), 2) ND1.2 (induced PSC), and 3) SFC840 (induced PSC); **macaque:** 1) MF12 (embryonic PSC) and 2-3) MF1 (embryonic PSC). RNA was taken from each line at ~D30.

Results of the Nanostring showed that all six cultures were expressing NPC-associated genes such as *VIM* and *PAX6* (**Figure 4.06a**). Welch's *t*-test found no significant difference in the expression of either *VIM* ($p = 0.48$) or *PAX6* ($p = 0.09$) between cultures of the two species. Of the 14 NPC genes examined, only one was significantly different between human and macaque inductions (*CCND2:* $p = 0.002$). Similarly, astrocytes, oligodendrocytes, and non-telencephalic cell types were extremely low in all six cultures (**Figure 4.06b**). They were also not significantly different between macaque and human ($p > 0.05$ for all, e.g. for the astrocyte-associated gene, *S100β*, $p = 0.35$).

Cortical genes like *EMX1* and *LHX2* were robustly detected in all cultures, indicating they were of cortical identity (**Figure 4.06c**). The macaque cultures appeared to express these genes more strongly than the human lines. The following genes were significantly higher in macaque than humans: *EMX1* ($p = 0.001$), *EMX2* ($p = 0.002$), *LHX2* ($p = 0.002$), *FEZF2* ($p = 0.003$), *NR4A2* ($p = 0.004$), and *SLC17A7* (p

[5] With thanks to Alessio Strano and Ellie Tuck of the Livesey lab.

= 0.005). This could be for one of two reasons: 1) the macaque cultures may contain a higher proportion of cortical cells, and/or 2) the macaque cultures may be more mature. In other words, there may be higher expression of cortical genes because more NPCs have differentiated – and the cultures may end up with comparable cortical identity later on. Given that macaque cortex is known to develop faster both *in vivo* and *in vitro* (4, 29), it is reasonable to deduce the latter reason may be occurring. However, the slightly elevated inhibitory content of the human cultures suggests the relative proportions of cell lineages may also be different (**Figure 4.06d**). That said, Welch's *t*-tests found only one inhibitory gene out of the 23 examined to be significantly elevated in macaque ($p = 0.001$).

Therefore, while there may be modest differences in either culture maturity or composition, the induction macaque and human inductions are overall reasonably similar.

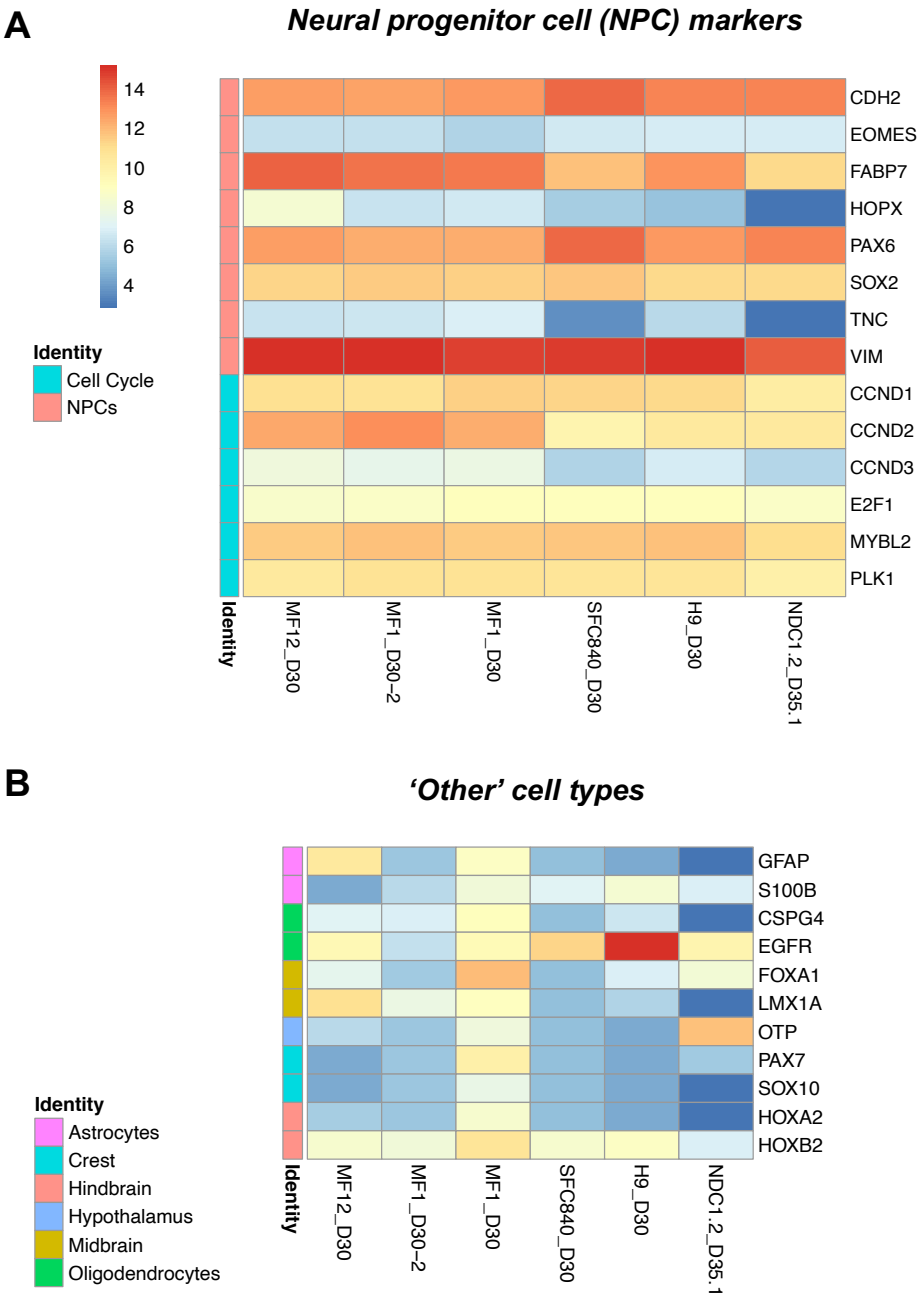


Figure 4.06. Transcriptional profiles of human and macaque inductions (figure 1 of 3)

Heatmaps show log2 normalized RNA counts of each gene measured by the Nanostring. Three inductions from each species were used with each sample representing 50ng total RNA. (A) Neural progenitor-associated markers are detected in all samples and to relatively equal levels across species. (B) Oligodendrocyte markers appear increased in the human inductions, while MF1 shows modest (but not significantly different) expression across several non-cortical cell types. Most cell types produced by non-cortical inductions are absent from the cultures.

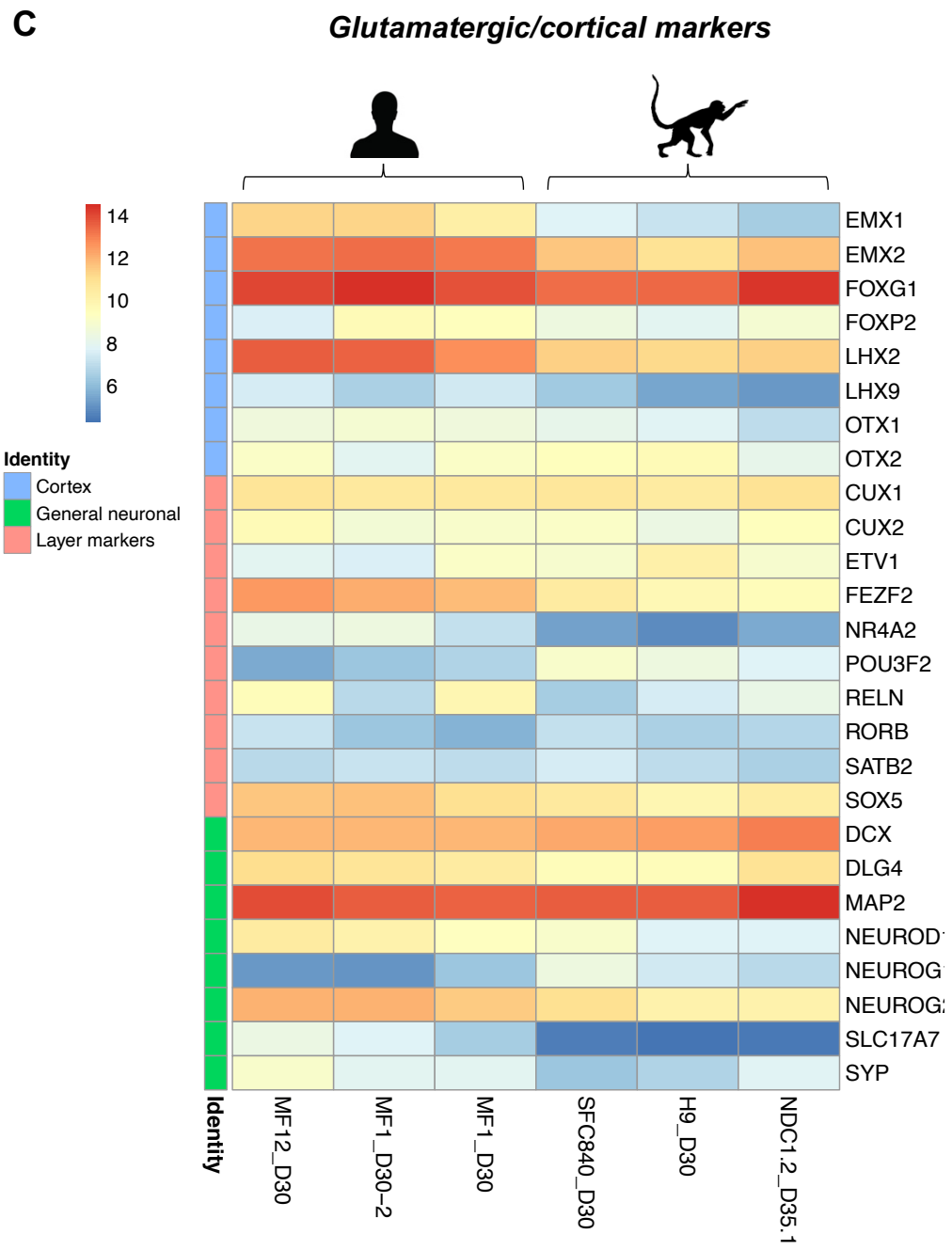


Figure 4.06. Transcriptional profiles of human and macaque inductions (figure 2 of 3)

Heatmaps show log2 normalized RNA counts of each gene measured by the Nanostring. Three inductions from each species were used with each time-point representing 50ng total RNA. Genes are grouped into three broad categories ("identities"): cortex-specific genes, general neuronal genes, and cortical layer markers. Cultures from both species express robust levels of cortical and neuronal markers. Descriptively, expression of several genes is slightly elevated in macaque – perhaps reflecting the faster development of the species.

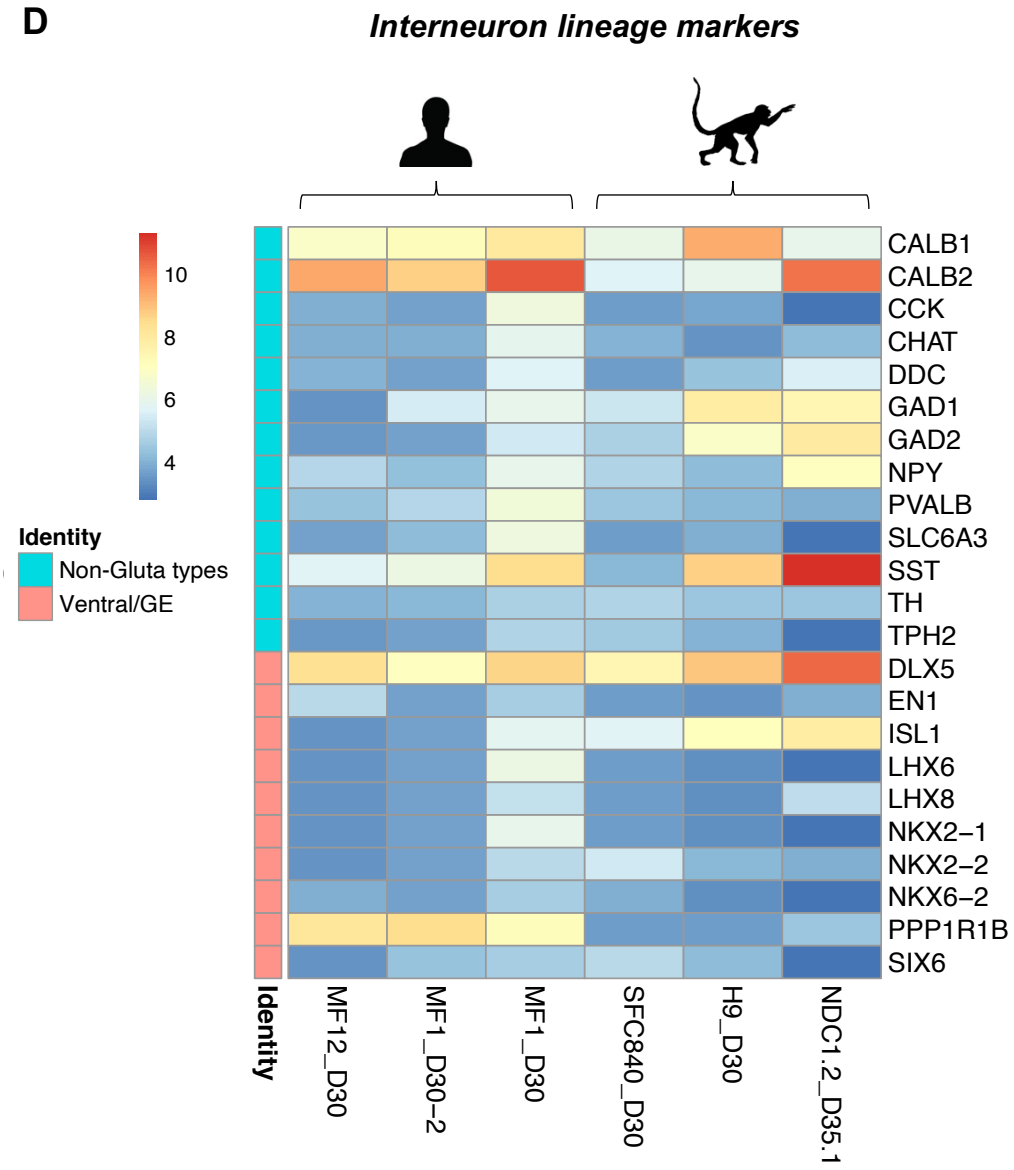


Figure 4.06. Transcriptional profiles of human and macaque inductions (figure 3 of 3)

Heatmaps show log2 normalized RNA counts of each gene measured by the Nanostring. Three inductions from each species were used with each time-point representing 50ng total RNA. Genes are grouped into two broad categories ("identities"): genes expressed in ganglionic eminence (GE; where cortical interneurons are formed), and GABAergic lineage markers (non-glutamatergic). Descriptively, the human inductions are slightly more ventralized (not statistically significant).

4.3.1 CNTNAP2 expression in human and macaque (mRNA)

Having confirmed both human and macaque inductions were comparable and cortical, I then proceeded to examine *CNTNAP2*/CASPR2 expression between species. RNA was collected at D50 and D80 from two human lines and two macaque lines (**Human:** H9 and NDC1.2, **Macaque:** MF1 and MF12) (one induction per line). I designed qPCR primers targeting two locations on the *CNTNAP2* gene. The first pair targeted exon 8 (which maps to only the canonical transcript CNTNAP2-201), and the second to exon 18 (which maps to protein-coding transcripts CNTNAP2-201 and CNTNAP2-207)(see **Figure 4.07**). Individual transcript-specific primers were tested, but had low efficiencies and were discarded for the (somewhat less specific for individual transcript isoforms) primers described above. All primer pairs had 100% sequence similarity between human and macaque. They were also tested for equal efficiency between species and the absence of off-target amplification.

A two-way ANOVA detected a significant interaction between species and culture age on *CNTNAP2* expression for both qRT-PCR experiments [**exon 8:** $F(1, 20) = 1.008$, $p = 3.27\text{e-}03$; **exon 18:** $F(1, 20) = 0.165$, $p = 6.89\text{e-}03$]. Analysis of simple main effects for species on *CNTNAP2* expression (Welch's ANOVA) reached significance after Bonferroni adjustment [**exon 8:** $F(1, 11.5) = 68.3$, $p < 0.000003$; **exon 18:** $F(1, 12.1) = 51.6$, $p < 0.00001$]. No significant difference was found for the effect of age on *CNTNAP2* expression [**exon 8:** $F(1, 21.7) = 0.18$, $p < 0.678$; **exon 18:** $F(1, 21.8) = 0.14$, $p < 0.715$]. These analyses were followed up by pairwise estimated marginal means comparisons. Analogous findings to the one-way ANOVA were detected: a significant increase in *CNTNAP2* expression is noted in humans relative to macaques at both timepoints [**exon 8 D50:** $p = 0.00005$, **exon 18 D50:** $p = 0.0001$, **exon 8 D80:** $p = 0.000002$, **exon 18 D80:** $p = 0.00004$]. No significant increase in expression is detected between timepoints (within species) for either primer pair.

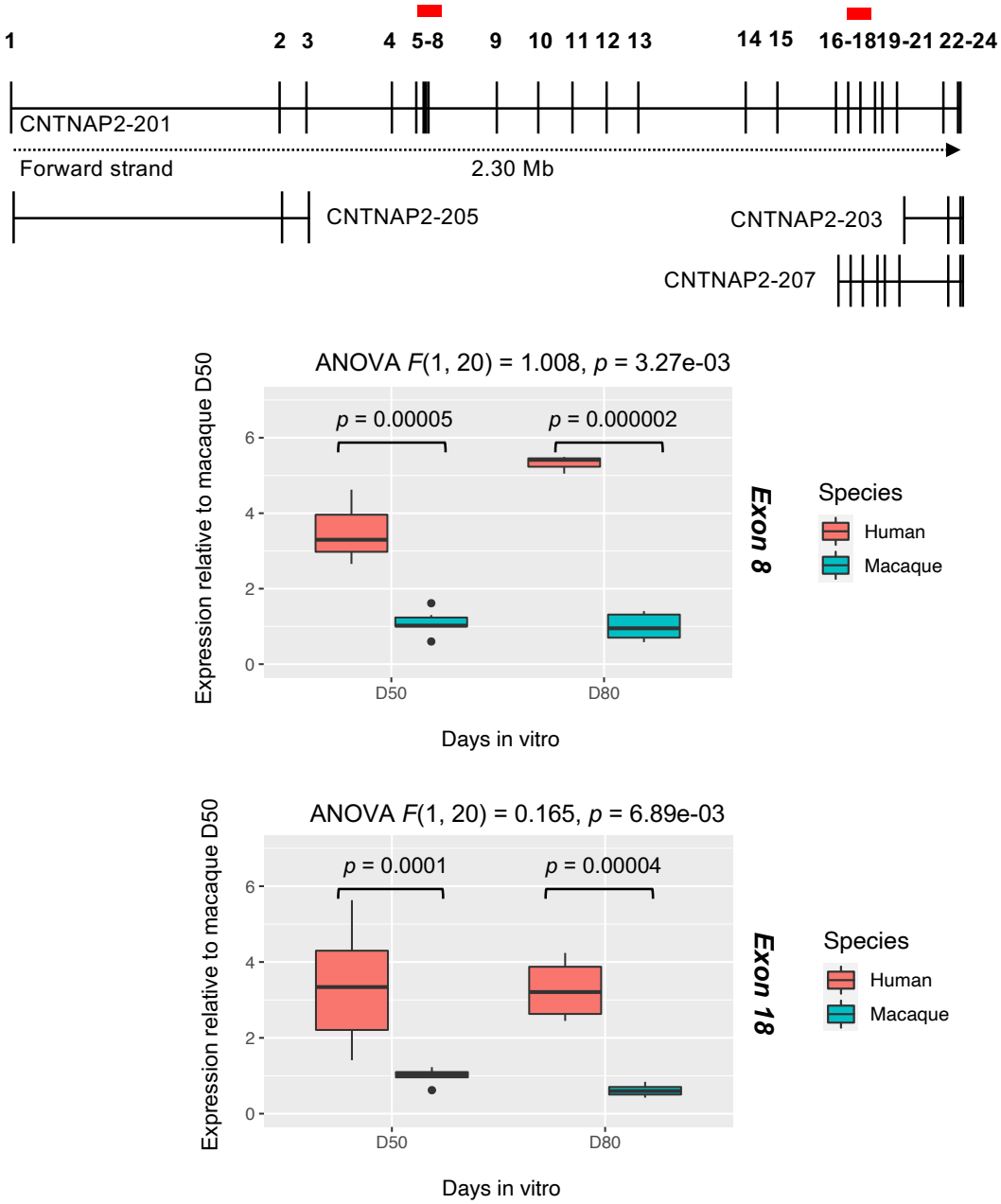


Figure 4.07. Differential expression of *CNTNAP2* (mRNA) between human and macaque in vitro cortical cultures

CNTNAP2 expression was measured by qRT-PCR from two human and two macaque lines (one induction per line, 3 samples per line). Expression was measured at two locations: exon 8 and exon 18 (where alternative transcript CNTNAP2-207 is also detected). Results were normalized to the geometric mean of two housekeepers, *PRPF6* and *DDB1*. At both regions of the *CNTNAP2* transcript – and at all timepoints - human *CNTNAP2* expression was significantly higher than macaque (two-way ANOVA followed by pairwise estimated marginal means analysis). Expression was not significantly different within species between timepoints.

4.3.2 CNTNAP2 expression in human and macaque (protein)

For analysis at the protein-level, I collected samples from the same human and macaque lines (**Human:** H9 and NDC1.2, **Macaque:** MF1 and MF12) (one induction per line). In addition to D50 and D80, I also collected protein from the NPC stage (D30). The same KO-validated CASPR2 antibody was used as previously described. This antibody, which binds to exons 20 – 23, shares 99% similarity with the macaque epitope (141/142 amino acids are conserved). The one change, which occurs at position 2 of the epitope, is a phenylalanine in humans and an isoleucine in macaques.

Crucially, the western blot experiments showed similar findings to what was observed with qRT-PCR. CASPR2 was consistently increased in human cultures across all timepoints (**Figure 4.08**). An independent samples *t*-test found these differences to be statistically significant [**D30:** $p = 0.003$; **D50:** $p = 0.03$; **D80:** $p = 0.00009$]. There was no visible CASPR2 protein band in the macaque D30 cultures, whereas there was clear expression for humans. By D50, the macaque samples expressed approximately one-quarter of the protein found in humans. Finally, by D80, expression in the macaque samples increased to approximately one-third of that in the human cells.

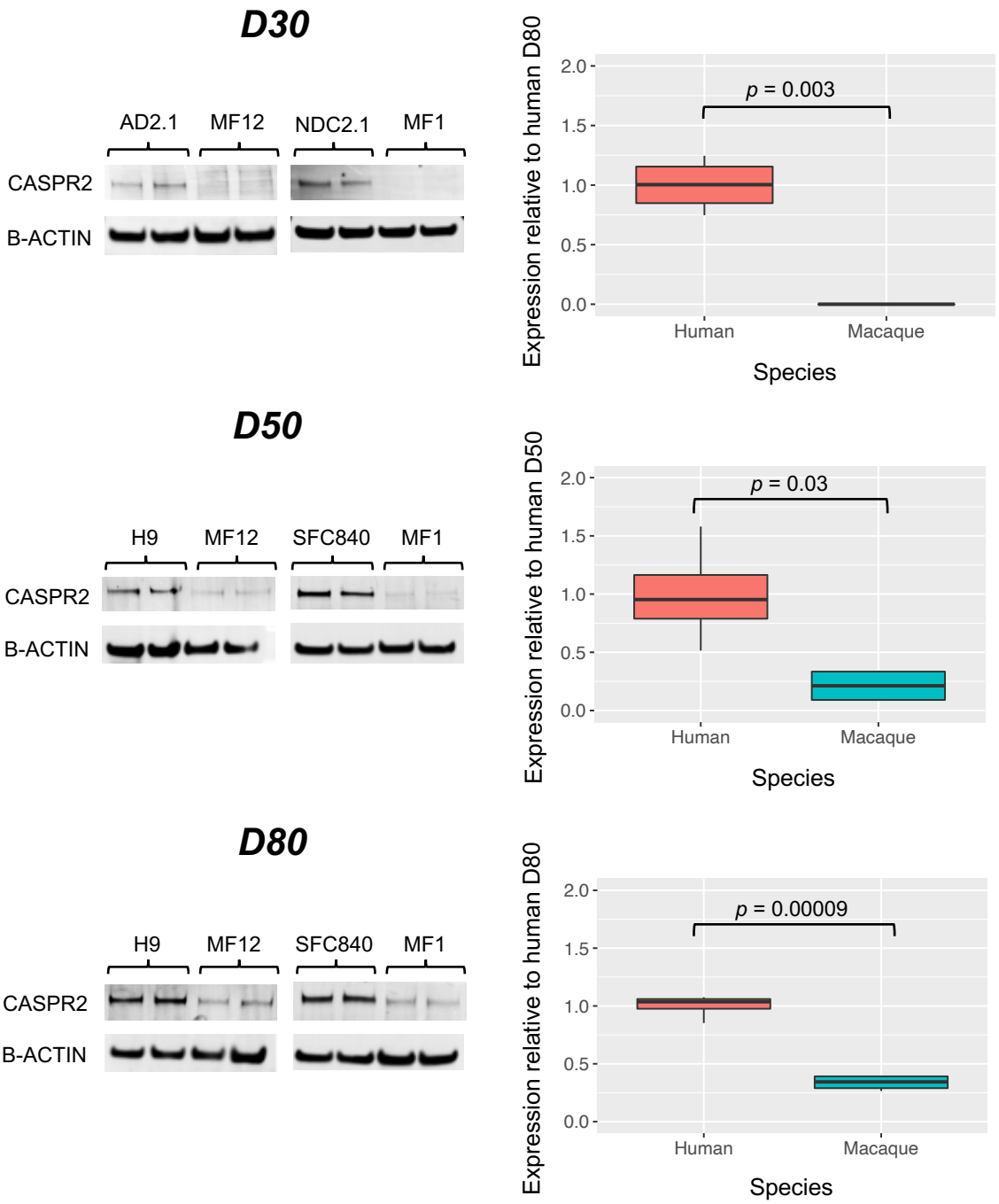


Figure 4.08. Differential expression of CASPR2 (protein) in human and macaque in vitro cortical cultures

CASPR2 expression was measured by western blot from two human and two macaque lines (one induction per line, 2 samples per line). Increased expression is noted in human cultures relative to macaque cultures at all three timepoints (Welch's t-test). Negligible expression is detected in macaque at D30. 40 µg protein was loaded per sample; data was normalized to B-ACTIN. The CASPR2 and B-ACTIN human antibodies had epitopes with >99% similarity in macaques, ensuring protein detection was as equal as possible between the two species.

4.4 The human *CNTNAP2* knockout line

As previously mentioned, mainly rodent and zebrafish *Cntnap2* KO lines currently exist (34, 36, 203, 245, 263, 269). Although tissue is available from patients with *CNTNAP2* loss-of-function mutations, these samples are subject to the limitations described in **Chapter 1**. For example, they cannot be experimentally manipulated, they are static/non-living tissues, and so on. Moreover, while PSC models have been generated from patient fibroblasts, there are currently only two models available. The first, was generated from a schizophrenic patient with a heterozygous *CNTNAP2* mutation (238, 239). The second, was generated from three patients with a c.3709DelG homozygous loss-of-function mutation (all three patients were diagnosed with cortical dysplasia focal epilepsy (CDFE)) (295). This existing KO model has so far only been used to generate forebrain organoids for RNA-Seq and analyses of organoid size. Additional human *CNTNAP2* KO studies are therefore needed – ideally involving a diverse set of loss-of-function mutations. Such models would be highly useful resources to experimentally study human *CNTNAP2* function. Considering the evidence that *CNTNAP2* dosage is important for human disease and evolution, determining what effects are caused a complete loss of *CNTNAP2* expression, will be an important first step.

With this in mind, I set out to create a human PSC *CNTNAP2* knockout line. These cells could then be differentiated with the Shi et al. (181) protocol to cortical neurons, or if desired, any other cell type with an existing differentiation protocol. To generate this line, I used CRISPR-Cas9, a revolutionary technique for genome engineering (296). CRISPR offers the ability to repair patient cell lines, or to introduce disease-causing mutations into control lines. This means isogenic cells that differ only in the variant(s) of interest can be compared. It also avoids the potential confounds that arise from using cells of different genetic backgrounds, and allows precise examination of only the candidate mutation(s). CRISPR protocols have been proven to be fast, easy, and highly effective (280). In-depth experiments can be performed to

study mechanisms underlying disease, and to test potential therapies. Finally, by targeting PSCs, other cell types that are difficult to manipulate – like post-mitotic cortical neurons – can be edited pre-differentiation.

4.4.1 CRISPR-Cas9 knockout design

A **Cas9-ribonucleoprotein (RNP) complex** was used to deliver a synthetic **guide RNA (gRNA)** to the 5' end of the canonical *CNTNAP2* transcript, *CNTNAP2-201* (**Figure 4.09**, strategy adapted from Bruntraeger et al. (280)). Although this gRNA does not target all isoforms of *CNTNAP2* (particularly those that cluster at the 3' end of the locus), it was chosen to increase the likelihood a non-functional gene product would be made of the canonical/most abundant transcript. Four gRNAs were tested targeting exons 2-4 (data not shown). Since only the gRNA targeting exon 3 was functional in the T7 assay (see **Section 4.4.2**), this gRNA was used CRISPR targeting. In order to efficiently screen for knockout clones, I designed a **homology directed repair template** to insert a **Pac1 restriction site** into the locus. Crucially, Pac1 sites contain an internal **TAA** stop signal - T**TAA**T^TAA. By restriction digest of DNA from targeted clones using Pac1 enzyme, cells containing the stop signal can easily be identified by multiple bands on a DNA gel (**Figure 4.12**). The repair template was designed as a **single-stranded DNA oligonucleotide (ssODN)**, which has been reported to improve insertion efficiency (280, 297). On either side of the Pac1 site were 50 bp **homology arms**; these target the stop codon to the cut site and were designed to ensure the stop was 'in-frame'. Care was also taken to remove the **protospacer adjacent motif (PAM)** to prevent repeat cutting by Cas9. All CRISPR editing reagents were delivered to PSCs via **nucleofection** (full details are reported in **Chapter 3**).

This RNP-ssODN strategy was chosen for several reasons: firstly (as mentioned), it is reported to have high cutting and editing efficiency - generating mutations in up to 98% of clones and inserting single nucleotide repairs in 40% of clones (280, 297). The editing reagents also degrade ~12-24 hours post-nucleofection, greatly reducing the chances of off-target mutations. Thirdly, HDR efficiency is shown to be higher with ssDNA templates compared to dsDNA (298-300). Similarly, non-specific integration of DNA is reduced since no dsDNA is introduced into the cells (280). Lastly, synthetic ssODNs and gRNAs can be ordered from many

commercial companies, avoiding the need to generate/validate targeting plasmids and/or lentiviral preps.

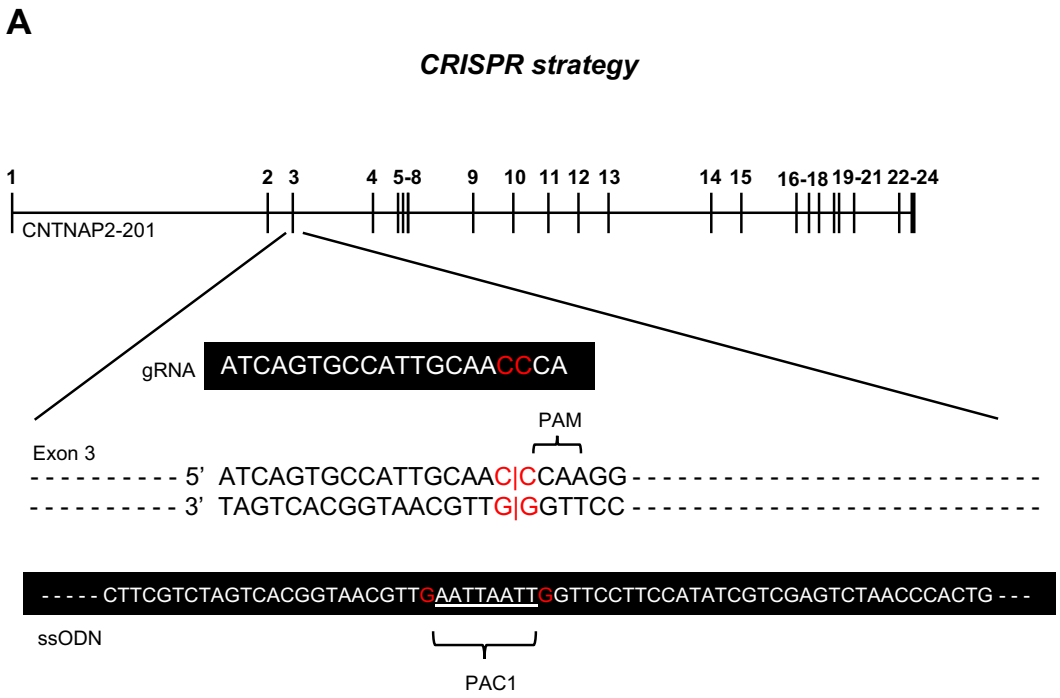


Figure 4.09. CRISPR-Cas9 *CNTNAP2* knockout strategy

(A) Guide RNA (gRNA) targeting exon 3 of the *CNTNAP2* gene. Cas9 cuts 3-4 nucleotides upstream of the protospacer adjacent motif (PAM), 'AGG'. The Cas9 cut site is highlighted in red along with corresponding nucleotides on the gRNA and single-stranded DNA oligonucleotide (ssODN) for reference. The ssODN acts a repair template to insert a Pac1 restriction site containing a 'TAA' stop codon. (B) Efficient editing results in the Pac1 restriction site signaling a premature stop (*). The resulting truncated protein will be destroyed by nonsense-mediated decay. The protein sequence shown corresponds to the fragment of exon 3 depicted above (starting with AAG = K).

4.4.2 CNTNAP2 KO validation

Post-nucleofection my first step was to check for signs of DNA editing. Bruntraeger et al. (280) reported ~20-25% of guides cannot edit due to low activity (even when used with functional Cas9). Although online tools can be used to predict gRNA efficiency, these predictions are not always accurate. To definitively test my *CNTNAP2* gRNA, therefore, I used a **T7 endonuclease assay** on a pooled population of nucleofected PSCs. This test works on the principle that T7 endonuclease cuts mismatched DNA. By PCR amplifying a region around the putative cut site, then denaturing and re-annealing the fragment in a thermocycler, DNA strands that contain Cas9-inflicted mutations may bind to wild type strands (forming a **'heteroduplex'**) (see **Figure 4.10**). The T7 nuclease will recognize and cut these heteroduplexes, which can then be identified as multiple bands on a DNA gel. In the event that no Cas9 cuts were made, only homoduplexes will form, and a single band will be present on the gel. As shown in **Figure 4.11**, testing of the *CNTNAP2* gRNA clearly showed the presence of multiple bands relative to the negative control (no T7 enzyme). With this confirmation that Cas9 editing had occurred, I then proceeded to isolate the PSCs into single clones for knockout screening.

Screening for Pac1/stop codon insertions were performed in a number of ways: 1) PAC1 digestion, 2) sanger sequencing, 3) MiSeq next generation sequencing, 4) qRT-PCR, and 5) western blot (**Figures 4.11-4.12**). In total, <70 clones were screened with 39 reading positive for the insertion of a PAC1 site (~56% targeting efficiency; data not shown). Sanger sequencing of the edited region confirmed at least 9 of 18 sequenced clones were positive. More importantly, due to the clear single peaks in the sequencing chromatograms, these clones appeared to be homozygous mutants. To definitively confirm whether clones were homozygous, heterozygous, or an accidental mix of multiple cells, two clones were sent for MiSeq next generation sequencing. The first clone had 4055/4140 reads (97.5% of all reads) detecting an eight nucleotide insertion, the exact size of the Pac1 insertion. The second clone had

4541/5172 reads (88% of all reads) detecting an eight base pair insertion. Clone 1 was therefore selected as a homozygous knockout, subject to confirmation that there was no *CNTNAP2* mRNA or protein expression.

Due to the low expression of *CNTNAP2* in PSCs and neural progenitor cells, RNA and protein samples were isolated from D50 neurons differentiated from the putative knockout PSC. The parental line, NDC1.2, was used as the WT control and was co-differentiated along with the KO cells. For the qRT-PCR, primer pairs were designed to measure both the 5' and 3' ends of the gene. Since two of the four *CNTNAP2* protein-coding transcripts cluster at the 3' end, but the chosen gRNA targeted the 5' end (to increase the chance a non-functional protein would be made), it is possible CNTNAP2-203 and CNTNAP2-207 were still being produced. Taking measurements at both the 5' and 3' ends could therefore help clarify whether this was occurring or not. As shown in **Figure 4.12b**, a dramatic reduction in *CNTNAP2* expression was detected via qRT-PCR. Pairwise comparisons with Welch's *t*-test determined this difference was significant for both primer pairs [**5'-end:** $p = 0.0001$, **3'-end:** $p = 0.008$]. Virtually no RNA was detected in the KO at the 5' end, and only ~15% of the WT expression was found at the 3' end.

For the protein level, I once again used my KO-validated antibody in a western blot. Since this antibody binds to the 3' end of the gene (exons 20-23), it would provide the ultimate evidence of whether the alternative transcripts were being expressed or not. Notably, the western showed a complete loss of CASPR2 protein (**Figure 4.12c**). Taken altogether, these results - at the gene, mRNA, and protein-level - suggest that this clone is a homozygous KO.

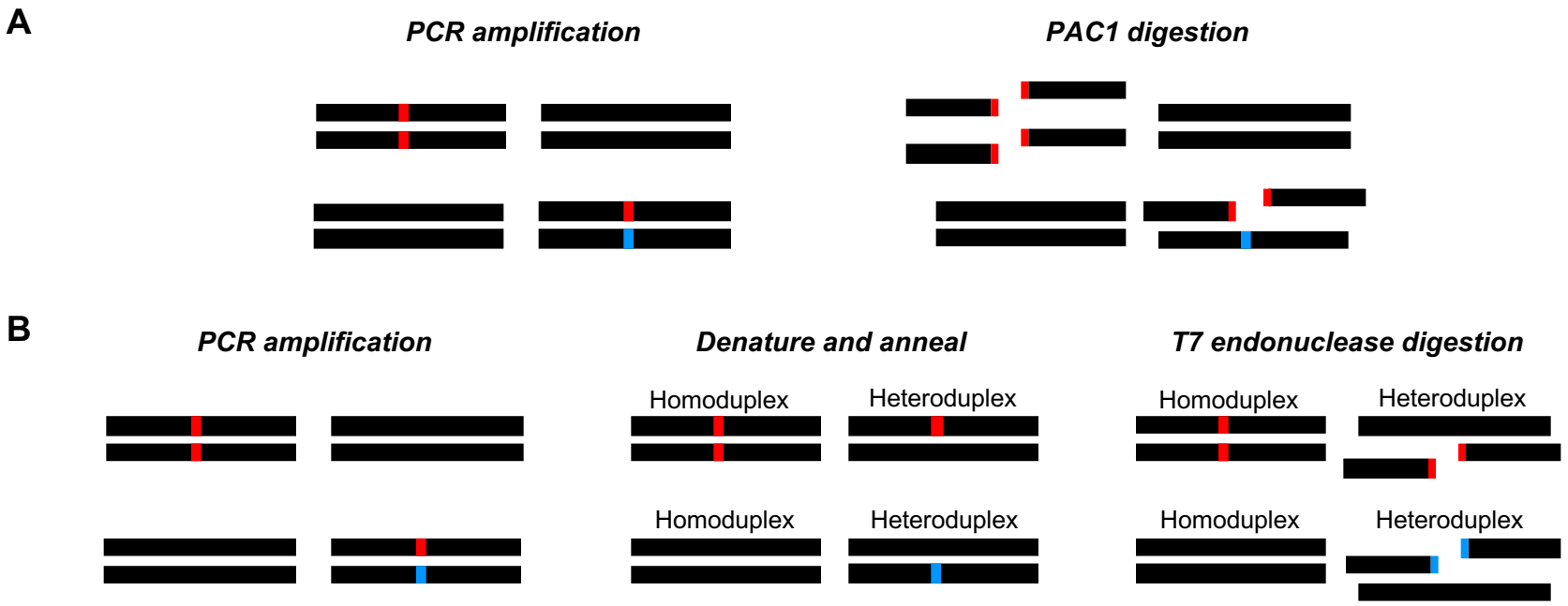


Figure 4.10. Schematic of the PAC1 and T7 endonuclease assays

(A) Black bars represent double stranded DNA. In the PAC1 assay, the editing site is amplified by PCR (with genomic DNA from each clone). The products are then incubated with PAC1 restriction enzyme at 37°C for one hour. If the PAC1 site (shown in red) has been successfully edited into the DNA, then multiple bands will be present on a DNA gel following digestion. DNA that has not been edited (all black) or DNA with a non-PAC1 mutation (blue) will not be digested. One band on a DNA gel could therefore mean no editing has occurred or Cas9 has inflicted a different mutation (e.g. indel, frameshift, etc.). (B) In the T7 endonuclease assay, a region around the editing site is amplified by PCR from pooled DNA (i.e. DNA taken from a mix of clones). The pooled DNA is then denatured and re-annealed in a thermocycler causing strands from different clones to come together. This process will form some homoduplexes (where both strands are not edited or contain the same edit) and heteroduplexes (where strands have been edited differently). As T7 endonuclease cleaves mismatched DNA, by incubating the samples with T7 enzyme, heteroduplex DNA will be cut – and identifiable by multiple bands on a DNA gel. The T7 endonuclease therefore shows whether mutant DNA is present in your original pool, signifying Cas9-mediated DNA editing has likely occurred.

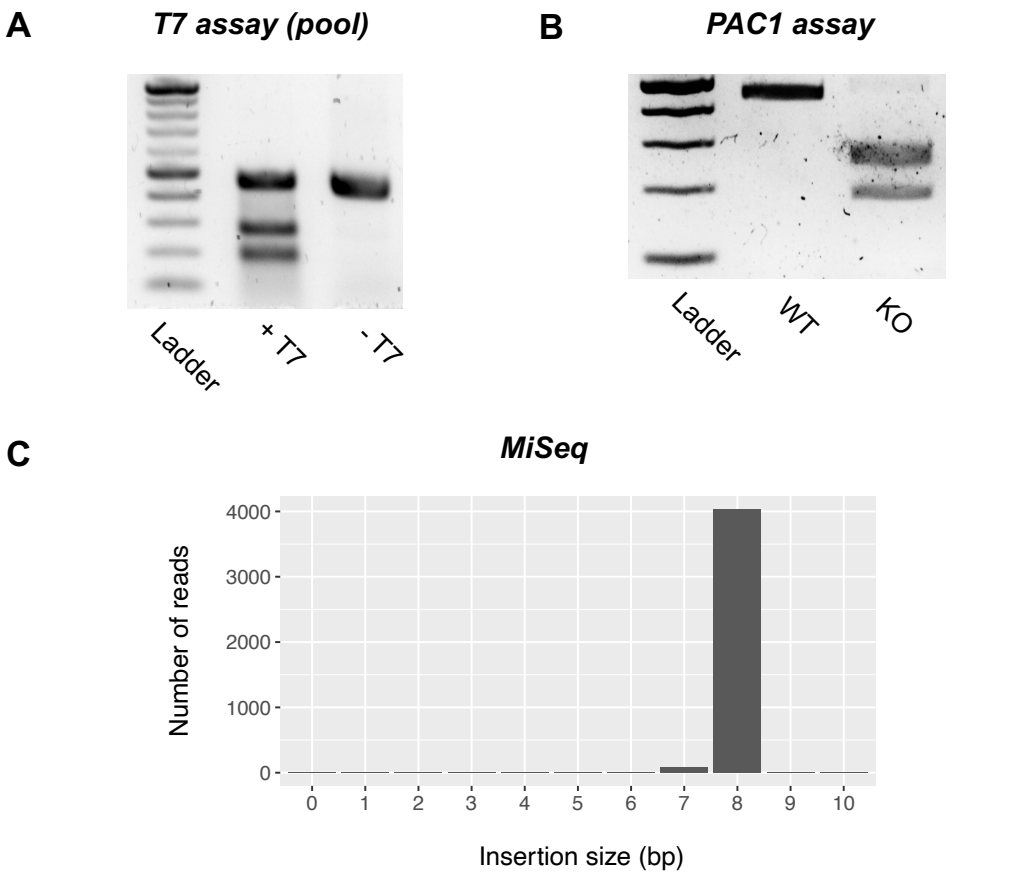


Figure 4.11. Knockout clone screening

(A) T7 endonuclease assay of pooled DNA taken from human PSCs targeted with the *CNTNAP2* gRNA. The presence of multiple bands indicates the amplified region contains mismatched/mutated DNA (relative to the control sample without T7 endonuclease). CRISPR editing is therefore likely to have occurred and the gRNA is likely to be effective. (B) PAC1 assay of a confirmed homozygous KO clone, showing two bands of the expected sizes (~190 bp and ~280 bp) compared to WT (~470 bp). (C) Next generation sequencing results of DNA from one clone (Illumina MiSeq). Of 4140 reads, 4055 (98%) detect an 8 bp insertion (the size of the PAC1 site), 78 reads detect a 7 bp insertion (1.8%), and 7 reads detect a 9 bp insertion (0.2%). These results suggest the clone is a homozygous KO. (D) Sanger sequencing of the KO clone confirms the 8 bp insertion is the PAC1 site (TTAAT^TAA).

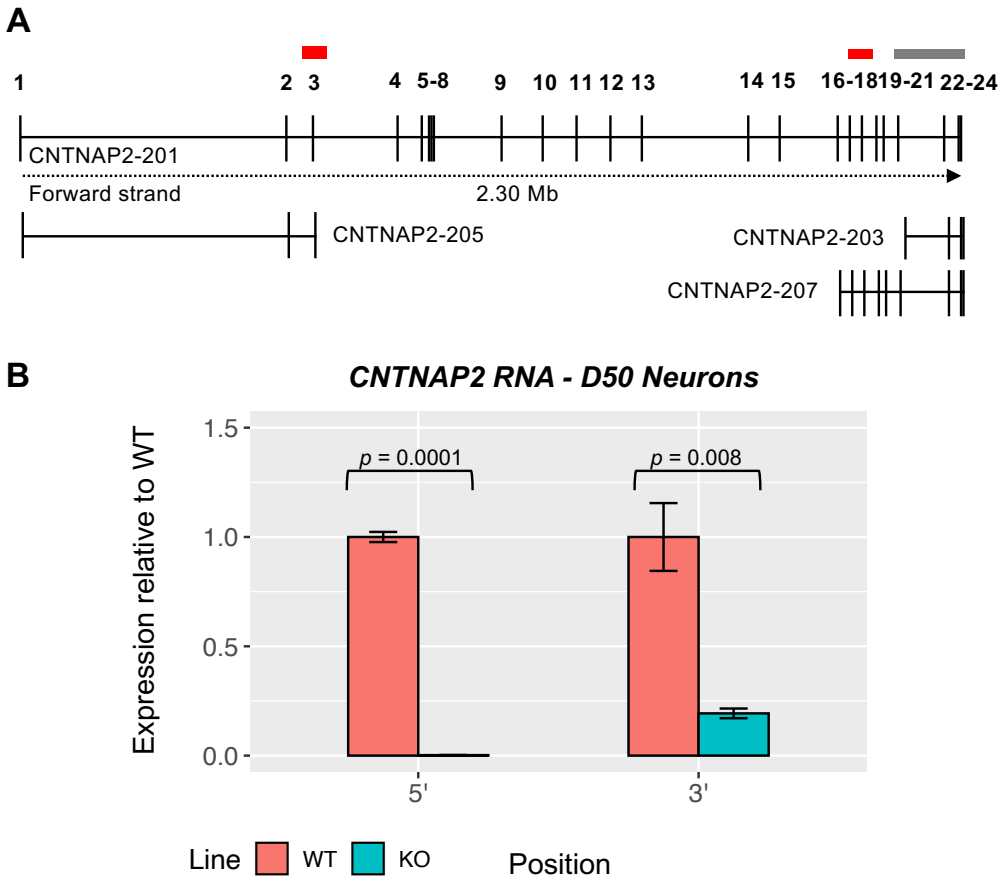


Figure 4.12. *CNTNAP2* mRNA and CASPR2 protein are reduced in the KO neurons (D50)

(A) Schematic of the four protein-coding *CNTNAP2* transcripts. The 5' qPCR primers (red) localize to exons 3-4 and detects transcript CNTNAP2-201. The 3' primers localize to exon 18 and detect CNTNAP2-201 and CNTNAP2-207. Finally, the CASPR2 antibody (grey) binds to exons 21-23 and detects CNTNAP2-201, CNTNAP2-203 and CNTNAP2-207. (B) qPCR results of D50 *CNTNAP2* KO and WT cortical neurons. Welch's t-test shows mRNA at both the 5' end 3' ends of the locus are significantly reduced/absent in the KO. (C) Western blot of D50 cultures shows a complete absence of CASPR2 protein in the KO, but robust expression in the WT. These findings serve as additional confirmation that the identified clone was a full homozygous KO.

4.4.3 Nanostring of the CNTNAP2 KO forebrain cultures

As with the human and macaque cultures, I used the Nanostring to check my KO and WT cultures were cortical and comparable to each other. More importantly, I wanted to see if *CNTNAP2* was required for cortical induction. To investigate, I collected RNA samples from two WT and two KO inductions (each at D50). Results from the Nanostring showed that these four cultures were highly similar to each other. No significant differences in gene expression were detected between WT and KO lines ($p > 0.05$, multiple pairwise comparisons with Bonferroni correction). Each were expressing cortical markers like *EMX2* and *LHX2* (**Figure 4.13**). Welch's *t*-test confirmed there was no significant difference in cortical gene expression between the two genotypes (e.g. for *EMX2,* $p = 0.51$). Similarly, NPC markers like *VIM* and *PAX6* were robustly expressed by all four inductions, and not significantly different from each other [***VIM:*** $p = 0.38$, ***PAX6:*** $p = 0.58$].

Inhibitory genes like *GAD2* were weakly expressed by all cultures, indicating few GABAergic cells are likely to be present. There was also no significant difference in inhibitory gene expression between genotypes (e.g. for *GAD2* $p = 0.43$). Moreover, non-telencephalic cell types like midbrain and hindbrain were absent from all cultures. Glia-associated genes (e.g. astrocyte and oligodendrocyte genes) were present yet weakly expressed. For example, the oligodendrocyte-associated gene *OLIG2* was faintly detected by the Nanostring (log2 counts = ~6 for both genotypes). No significant difference in expression was found between WT and KO ($p = 0.83$).

Overall, these results suggest the inductions between WT and KO are similar in composition, maturity, and quality. It also suggests that *CNTNAP2* is not needed for correct cortical differentiation. A more detailed transcriptomic comparison of WT and KO cells will be carried out with scRNA-Seq in the next chapter.

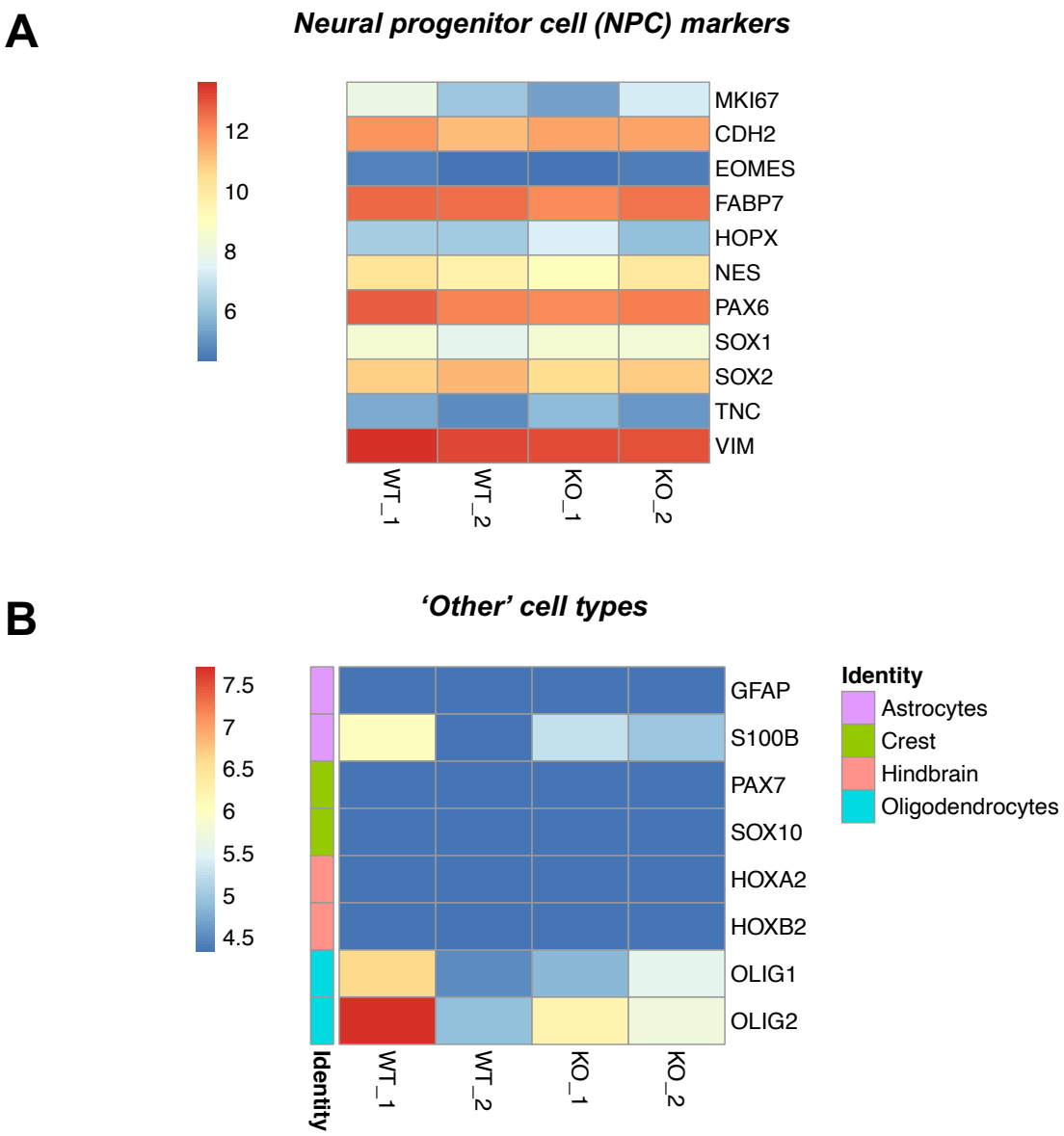


Figure 4.13. Transcriptional profiles of *CNTNAP2* WT and KO cortical inductions (figure 1 of 3)

Heatmaps show log2 normalized RNA counts of each gene measured by the Nanostring. Two D50 inductions from WT and two D50 inductions from KO were used. Each sample represents 50ng total RNA. (A) Neural progenitor-associated markers are detected in all samples and to relatively equal levels across genotypes. (B) Most cell types produced by non-cortical inductions are absent from the cultures. Astrocyte genes (*GFAP* and *S100B*) and oligodendrocyte genes (*OLIG1* and *OLIG2*) are not significantly different between the cultures. That said, WT_1 has a modest increase in *OLIG1/2* (note the scale bar exaggerates the difference).

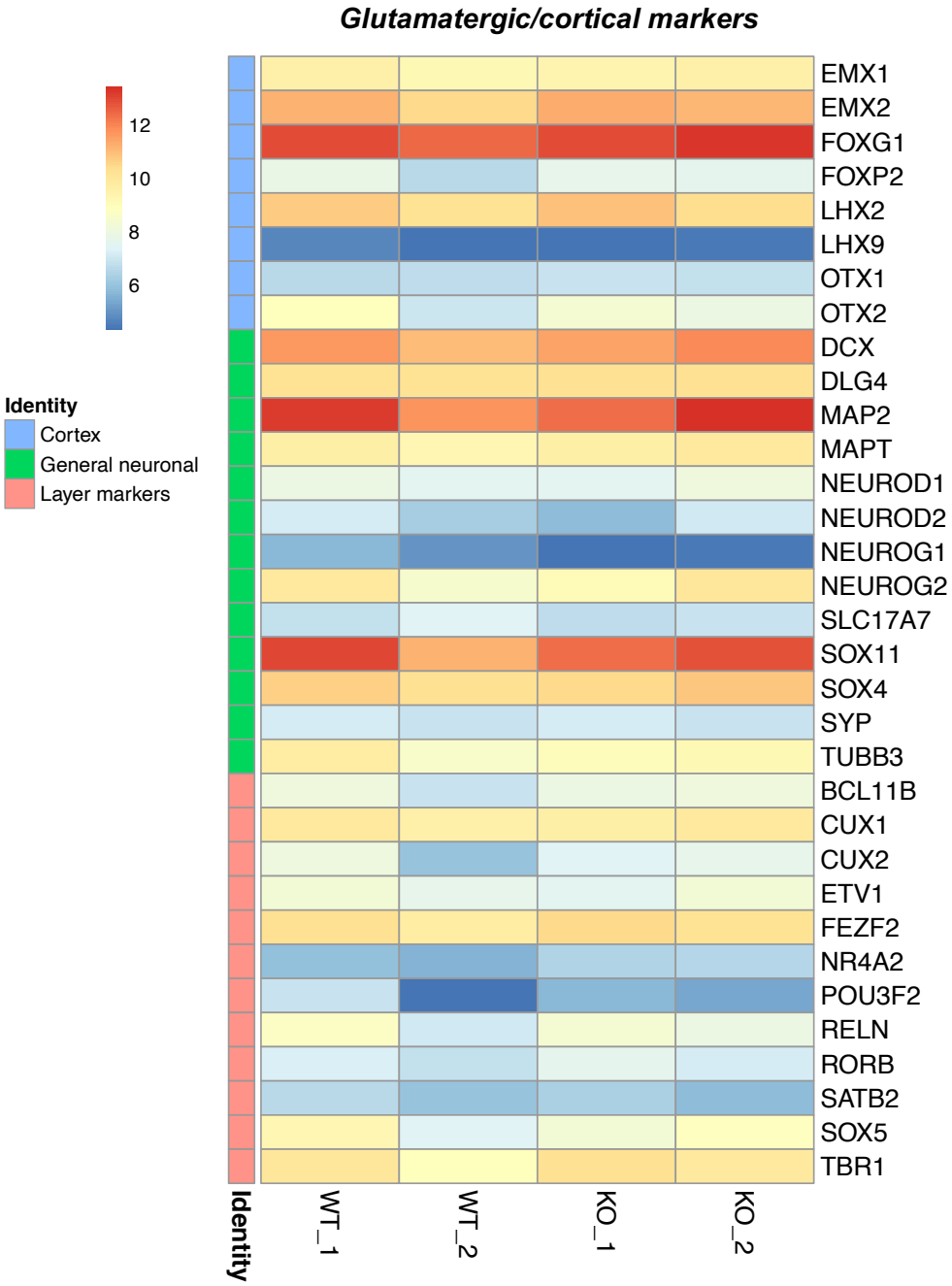


Figure 4.13. Transcriptional profiles of *CNTNAP2* WT and KO cortical inductions (figure 2 of 3)

Heatmap showing log2 normalized RNA counts of each gene measured by the Nanostring. Two D50 inductions from WT and two D50 inductions from KO were used. Each sample represents 50ng total RNA. Genes are grouped into three broad categories ("identities"): cortex-specific genes, general neuronal genes, and cortical layer markers. Cultures from both genotypes express robust levels of cortical and neuronal markers.

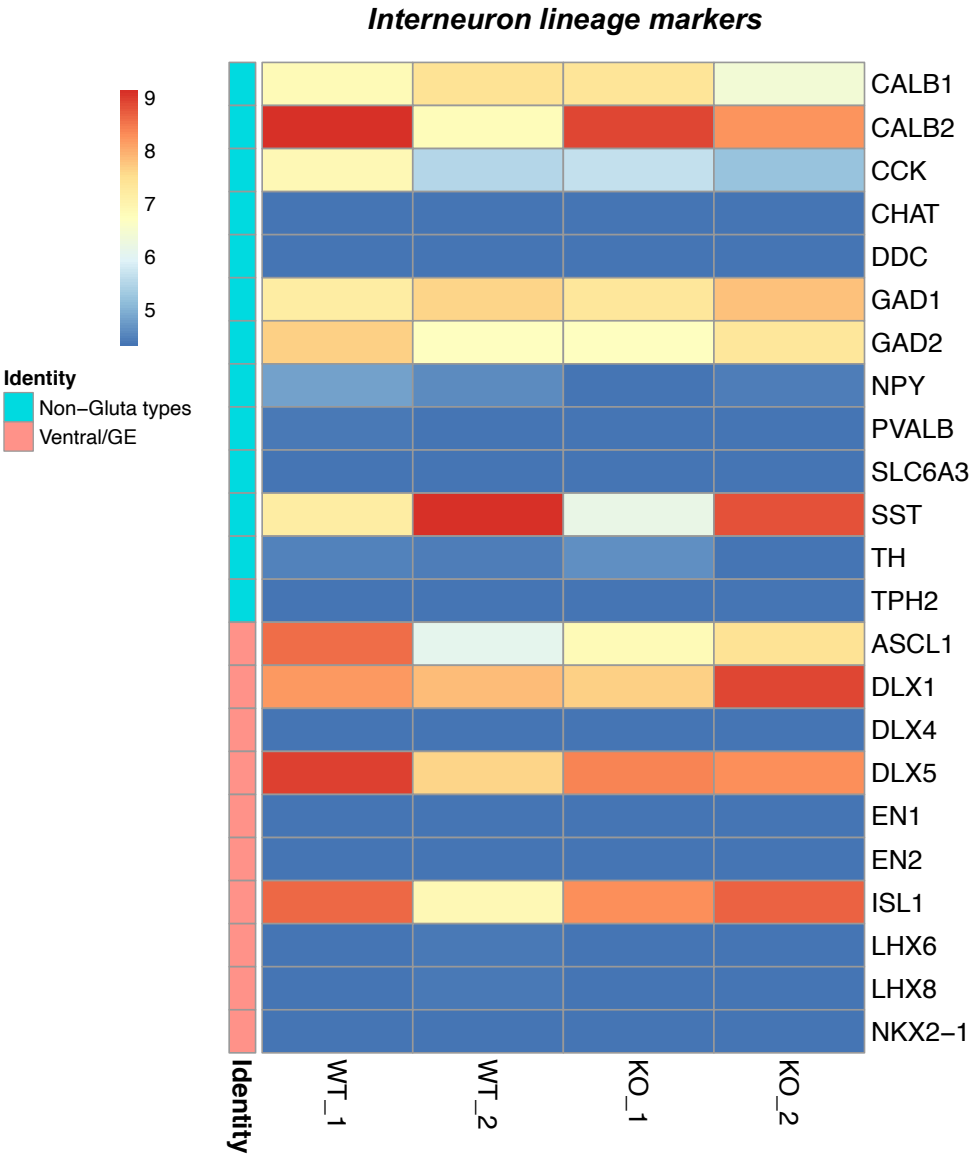


Figure 4.13. Transcriptional profiles of *CNTNAP2* WT and KO cortical inductions (figure 3 of 3)

Heatmap showing log2 normalized RNA counts of each gene measured by the Nanostring. Two D50 inductions from WT and two D50 inductions from KO were used. Each sample represents 50ng total RNA. Genes are grouped into two broad categories ("identities"): genes expressed in ganglionic eminence (GE; where cortical interneurons are formed), and GABAergic lineage markers (non-glutamatergic). Overall, there is no strong difference in inhibitory gene expression between the two genotypes. All cultures are also relatively low in inhibitory gene expression (with a few exceptions such as *DLX5*).

4.5 Discussion

In this chapter, I have shown that human *CNTNAP2* expression increases over *in vitro* cortical development (**Figures 4.03-4.04**). Starting from no expression in stem cells, *CNTNAP2* becomes stably expressed by ~D50. Secondly, I confirmed that CASPR2 is expressed in synaptic puncta on axons, dendrites, and cell bodies (similar to the expression previously described for mouse cortex (34, 202, 203)) (**Figure 4.05**). Thirdly, I detected *CNTNAP2* expression in human PSC-derived forebrain neurons was higher than in macaque PSC-derived cultures - across D30, D50, and D80 timepoints (**Figures 4.07-4.08**). Finally, I introduced the second human *CNTNAP2* KO cell line (to my knowledge), and described the steps involved in generating and validating the KO (**Figures 4.09-4.12**).

Published scRNA-seq experiments of human fetal cortex show *CNTNAP2* is expressed by NPCs and deep layer excitatory neurons (post-conception weeks 6-37) (287). Our data agree with these findings as expression is noted from ~D30, when these cell types are present in the culture. It also follows that *CNTNAP2* levels would increase with developmental time, as more cells and cell types expressing the gene become abundant (e.g. deep/upper layer excitatory neurons and interneurons). The highest expression of *CNTNAP2* has previously been reported in layers II-III of the frontal and temporal cortex. Since these are later-born cell types in our system (generated by ~D80), this would also support an increase in expression with time (195, 196).

The finding that *CNTNAP2* expression increases in our neuronal cultures by ~D50 is interesting for several reasons. *CNTNAP2* has been repeatedly implicated in neurite outgrowth (35, 202, 261), dendritic spine formation (34-36, 263), and in the formation of synaptic networks (35, 194, 203, 245, 257, 262, 263, 271, 301) (see **Chapter 2**). These processes begin around D50 in Shi et al. cultures (173, 181, 186), meaning there could be a connection between the appearance of *CNTNAP2* and their development. More detailed discussion of *CNTNAP2* expression at the single cell

resolution can be found in **Chapter 5**. Further investigation into the connection between *CNTNAP2* and synaptic function can be found in **Chapter 6**.

Regarding the spatial location of CASPR2, super resolution microscopy has previously shown the protein is localized to both excitatory synapses (PSD-95$^+$/vGLUT1$^+$) and inhibitory synapses (gephyrin$^+$/vGAT$^+$) in mouse cortex (203). However, CASPR2 was found to be more abundant in inhibitory synapses - detected in 61% versus only 45% of excitatory synapses. While this was not covered in my experiments, it would certainly be worthwhile investigating whether this distribution is also found in human cortex. It would also be useful to clarify whether CASPR2 puncta are mainly found in the pre-synaptic or the post-synaptic compartments. While my dataset was too small to extract any meaningful conclusions about this, discerning which part of the synapse CASPR2 is located to – and whether it is primarily found at excitatory or inhibitory synapses - will be important for fully elucidating its function/mechanism of action.

One of the most striking results from this chapter was the finding that our human cortical neurons have much stronger expression of *CNTNAP2* than the macaque neurons – about a 3-fold increase. A caveat to this finding was that the antibody used for the western blots had a one amino acid difference in the macaque epitope. In theory, this could affect antibody binding efficiency, and thus lead to a weaker signal. However, given that only a single amino acid (out of 142) did not match, it would be surprising for this to cause such a marked in antibody binding. Moreover, my qRT-PCR data – which was measured with two separate primer pairs that had 100% sequence similarity/equal efficiency between the two species – detected the same ~3-fold difference. This observation is therefore unlikely to be a false positive. I was unable to find another antibody with higher epitope similarity in macaques, however, should one become available this would be a worthwhile follow-up study to confirm the findings of this experiment.

It is also possible that the increase in *CNTNAP2* expression in humans versus macaque cultures was caused by differences in culture composition. For example, if there was a higher proportion of cells in the macaque cultures that did not express *CNTNAP2*, this could dilute the overall readings. However, the Nanostring data showed the cultures were not significantly different from each other. One would also expect that if *CNTNAP2* expression increases with culture maturity (as shown), and considering that macaque cortex develops faster (4, 29), that macaques would have an increased expression and not humans.

The finding that human cortex had higher *CNTNAP2* expression than macaques also agrees with previous findings in the field. As mentioned, Nowick and colleagues (37) found a 1.3-fold increase in *CNTNAP2* expression in adult human pre-frontal cortex compared to chimpanzees. This finding was then corroborated by Muntane et al. (39), who also showed increased expression on the human-chimpanzee lineage (in comparison to Gibbons, Old and New World Monkeys, Lemurs, and Lorises). However, it does not agree with a scRNA-Seq study by Pollen et al. (3) who found no difference in *CNTNAP2* expression between human and macaque primary fetal cortex (post-conception weeks 9-22). In PSC-derived cerebral organoids, the authors also found *CNTNAP2* was more highly expressed in chimpanzee neurons than in human neurons (i.e. the opposite direction). They did not generate macaque organoids, which otherwise would have been a useful comparison to make to our experiment. That said, one reason for the discrepancy between the Pollen data and my findings could be the scale of the respective experiments. In particular, the Pollen experiment was a genome-wide study that was aimed to inform on general trends across multiple genes – and not one particular gene. It is possible that this meant their experiment was not optimized to detect *CNTNAP2* in particular. As such, their findings may not be as sensitive or as accurate. It is also possible that our *in vitro* experiment leads to different results than observed *in vivo*, or that our bulk approach (versus their single cell approach) is creating the discrepancy.

With all this in mind, there are a few important next steps to take. Ideally, our *in vitro* studies should be complemented by work in primary tissue. While PSC-derived models are largely considered to be transcriptionally similar to their primary counterparts (289), verifying one's findings in 'true' cortical tissue is nevertheless important. So long as PSC models do not share 100% identity with primary tissue, there are a subset of genes whose transcription differs between *in vitro* and primary cells. Consequently, its critical to verify that *CNTNAP2* is not one of them. Studying a larger number of non-human primate species will also be a useful next step. Including chimpanzees or bonobos will be needed to understand *CNTNAP2* expression differences between our closest primate relatives. A difference in expression between humans and macaques is interesting on its own, however, a difference in expression between chimpanzees/bonobos would be even more suggestive that *CNTNAP2* in important for human evolution. While existing studies have preliminarily shown this is the case in primary adult cortical tissue, these experiments are few in number. Replication by independent scientists will therefore be crucial.

In addition to analyzing a greater number of primate species, it would also be useful to examine expression over a greater number of developmental stages. While access to non-human primate fetal samples is restricted, simply increasing the range of ages (e.g. juvenile, adult, elderly) would still be informative. Considering our *in vitro* study, as shown in the Nanostring data, by D80 our cultures were still somewhat immature (e.g. they lacked upper layer neuron markers and glia). Again, considering the evidence that *CNTNAP2* is involved in synaptic transmission, and given that these processes develop at the later stages of corticogenesis, looking at expression changes at later timepoints will be important – for both comparisons of humans and other primates, but also for understanding the temporospatial changes to *CNTNAP2* in humans alone.

The final section of this chapter introduced our *CNTNAP2* KO PSC line. Functional studies on cortical cultures derived from these cells will be performed in **Chapter 6**. For now, it is simply worth mentioning a few possible next steps involving

the line. Firstly, validating there were no off-target effects (e.g. by copy number variation (CNV) analysis or genome sequencing) will be important to ensure any observed phenotypes are down to loss of *CNTNAP2* only. It would also be worthwhile creating lines that recapitulate specific patient mutations. It would also be informative to try and correct these mutations with CRISPR. If the phenotypes are able to be reversed, then gene therapies could be a potential way to treat patients with *CNTNAP2* mutations. As stated, there is currently only two PSCs line from *CNTNAP2* patients (238, 239, 295). In **Chapter 2**, I discussed how mutations in *CNTNAP2* can be differently penetrant between patients (43). Some patients who have heterozygous mutations may have more severe disease than others, and it is not currently clear why. Investigating the mechanisms by which different heterozygous mutations cause different phenotypes, and why certain carrier parents are unaffected, will be immensely important to treating illnesses caused by *CNTNAP2* mutations. Finally, generating additional *CNTNAP2* KOs in other cell lines (e.g. distinct genetic backgrounds), and by targeting other parts of the gene will also be required next steps. This would help confirm that any observed phenotypes in our current KO model are generalizable.

Chapter 5

Single cell RNA sequencing studies of CNTNAP2 WT and KO PSC-derived human cortical cultures

5.1 Introduction

Since its inception in 2009 (302), **single cell RNA-sequencing (scRNA-Seq)** studies have revolutionized our ability to i) identify and classify cell types, ii) characterize rare or small cell populations, and iii) follow cells along dynamic processes, such as differentiation, development, or disease progression (303-305). Perhaps most importantly, scRNA-Seq allows cell type-specific expression to be resolved from a diverse population of cells. Given the heterogeneity and complexity of the brain (and

it's *in vitro* models (3)), rare cell types and transcripts are common (306). These are therefore at risk of being diluted out in bulk experiments.

In the previous chapter, I examined bulk *CNTNAP2* expression in our PSC-derived forebrain cultures. Using qRT-PCR and western blot assays, I detected that *CNTNAP2* mRNA and protein increased in expression with culture development. I suggested this was because *CNTNAP2* expression is strongest in cortical neurons (rather than progenitors), which increasingly populate the Shi et al. (181) cultures with time (according to the Nanostring data). However, without single cell resolution it was not possible to definitively say which cell types were expressing *CNTNAP2*, or their relative strengths of expression. As such, performing scRNA-Seq on our PSC-derived forebrain cultures would be highly informative. Not only would it pinpoint which cell types express *CNTNAP2* (and how much or how little), but it would also inform on how *CNTNAP2* expression changes across developmental time. For example, it was not possible to distinguish whether the low *CNTNAP2* expression noted at D30 was due to high expression in a small number of neurons[6], or low expression in many progenitors (which should be common at D30).

To my knowledge there is also no existing scRNA-Seq study of a human *CNTNAP2* KO forebrain system. Therefore, in addition to analysing WT cultures, I also made use of the *CNTNAP2* KO line introduced in the previous chapter. This would greatly inform on the genes and pathways that get dysregulated by a loss of *CNTNAP2*. It would also highlight cell types that are particularly vulnerable to such losses. In turn, this knowledge could provide significant insights into the mechanisms underlying *CNTNAP2* loss-of-function phenotypes (e.g. debilitating diseases such as **cortical dysplasia focal epilepsy (CDFE)** and **Pitt-Hopkins syndrome (PTHS)** (189)). It would also shed light on the processes dysregulated by heterozygous *CNTNAP2* mutations (e.g. autism, schizophrenia, intellectual disability) (43). Finally, the effects of

[6] A small number as neurons are rare at D30.

CNTNAP2 KO in cells that do not express the gene can also be investigated (i.e. both cell autonomous and non-cell-autonomous effects).

Initial published data suggests a scRNA-Seq comparison of WT and *CNTNAP2* KO human neurons could be highly interesting. Firstly, Flaherty et al. (239) conducted bulk RNA-Seq on cortical neurons containing a heterozygous *CNTNAP2* mutation. These neurons were derived from PSCs generated from a patient with schizophrenia. The authors identified **differentially expressed (DE)** genes enriched in processes such as DNA binding, central nervous system (CNS) development, and synaptogenesis. A second RNA-Seq study was performed on PSC-derived forebrain organoids with a *CNTNAP2* homozygous loss-of-function mutation (295). The organoids were generated from patients with cortical dysplasia focal epilepsy. The DE genes identified by this second group included genes involved in cell proliferation and neurogenesis. Crucially, there were also a significant number of autism-associated genes. A final RNA-Seq study was carried out on hippocampal tissue from *Cntnap2* KO mice (307). The DE genes identified here were enriched in pathways relating to axon guidance and PI3K-Akt signaling. A comparison of our scRNA-Seq findings with these existing results will be made in the discussion of this chapter.

Another area scRNA-Seq could shed light on is whether there are losses of certain cell types in *CNTNAP2* patient cortex. This would be particularly useful given the evidence in mice that *CNTNAP2* KO leads to a loss of cortical interneurons. Specifically, Peñagarikano et al. (245) discovered that there was a ~20% loss of parvalbumin positive (PV) interneurons (across all cortical layers), and ~10% losses of calretinin-positive (CALB2) and neuropeptide Y-positive (NPY) interneurons. Similar findings have since been reported in rodent and zebrafish models by Hali et al. (272), Selimbeyoglu et al. (271), and Hoffman et al. (269). The loss of these interneurons was hypothesized to underly the frequent seizures observed in the model organisms, and would also fall in line with an excitation : inhibition imbalance leading to CDFE/PTHS in humans. As of 2020, however, a reduction in cortical

interneurons has never been identified in humans. Clarifying whether there is such a reduction in a human model would be immensely useful to treat *CNTNAP2* patients.

With these points in mind, I set out to examine transcriptome-wide changes due to loss of human *CNTNAP2* in our PSC-derived system. **scRNA-Seq** was performed on two WT and two KO D50 inductions. This timepoint was chosen for several reasons. Firstly, it is at ~D50 that *CNTNAP2* expression is stably expressed in our cultures (see **Figures 4.03** and **4.04** in **Chapter 4**). Prior to this time, expression is weak and/or variable. As such, this period could represent when *CNTNAP2* is first required by cortical neurons. *CNTNAP2* has also been repeatedly implicated in neurite outgrowth (35, 202, 261), dendritic spine formation (34-36, 263), and in the formation of synaptic networks (35, 194, 203, 245, 257, 262, 263, 271, 301). These processes also begin around D50 in our cultures (173, 181, 186), meaning such a timepoint could inform on the role of *CNTNAP2* in their initial development.

5.2 Overview of single cell RNA sequencing

A notable component of our scRNA-Seq strategy was that the WT and KO cultures were **multiplexed** using **cell hashing antibodies** (see **Figure 5.01**). In cell hashing, separate samples are stained with distinct oligonucleotide-tagged antibodies (i.e. one antibody per sample) (281). Hashtag antibodies - referred to as **hashtag oligonucleotides**, or **HTOs** - target ubiquitously expressed cell surface proteins (CD298 and β2 microglobulin) (281). As such, they will bind to all cells in a sample irrespective of cell type. Staining the samples allows them to be pooled ahead of library construction, which reduces experimental costs, resources usage, and time. More importantly, it diminishes 'batch effects' that can introduce noise and obscure gene expression analyses (281). Finally, multiplexing enables the detection of **cell multiplets**, two or more cells that have been mistaken as one, by highlighting 'cells' that are tagged by more than one antibody. As more samples are pooled, the likelihood of a multiplet containing differentially labelled cells increases.

For our scRNA-Seq experiment, library construction was completed using the 10X Genomics Single Cell 3' kit (full details reported in **Chapter 3**)(see **Figure 5.01**). A 100-cycle run was performed on an Illumina NovaSeq to obtain approximately 800 million reads across the pooled samples. Overall, 18,304 cells were sequenced, with an average of 51,780 reads per cell mapping to a median of 1,818 genes per cell.

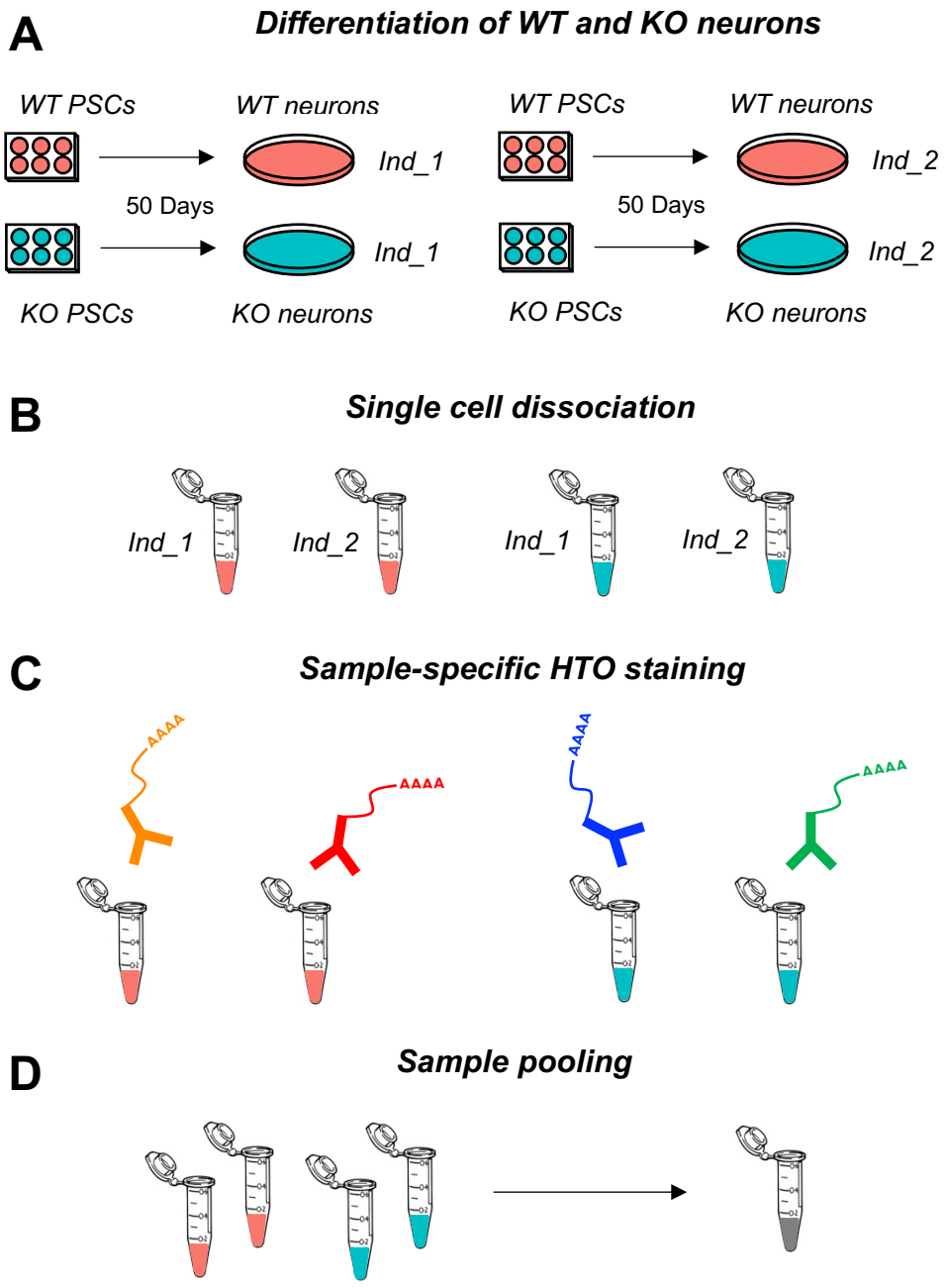


Figure 5.01. Overview of scRNA-Seq experiment

(A) Two independent inductions of WT and *CNTNAP2* KO cortical cultures were generated from PSCs and matured to D50. (B) At D50 the cultures were dissociated into single cell suspensions. (C) Each sample (i.e. WT induction_1, KO induction_1) was stained with a unique hashtag oligonucleotide (HTO). HTOs are antibodies that allow multiple samples to be pooled together before cell encapsulation (D). Such 'sample multiplexing' greatly reduces batch effects/noise in downstream gene expression analyses. *Figure continued on the next page.*

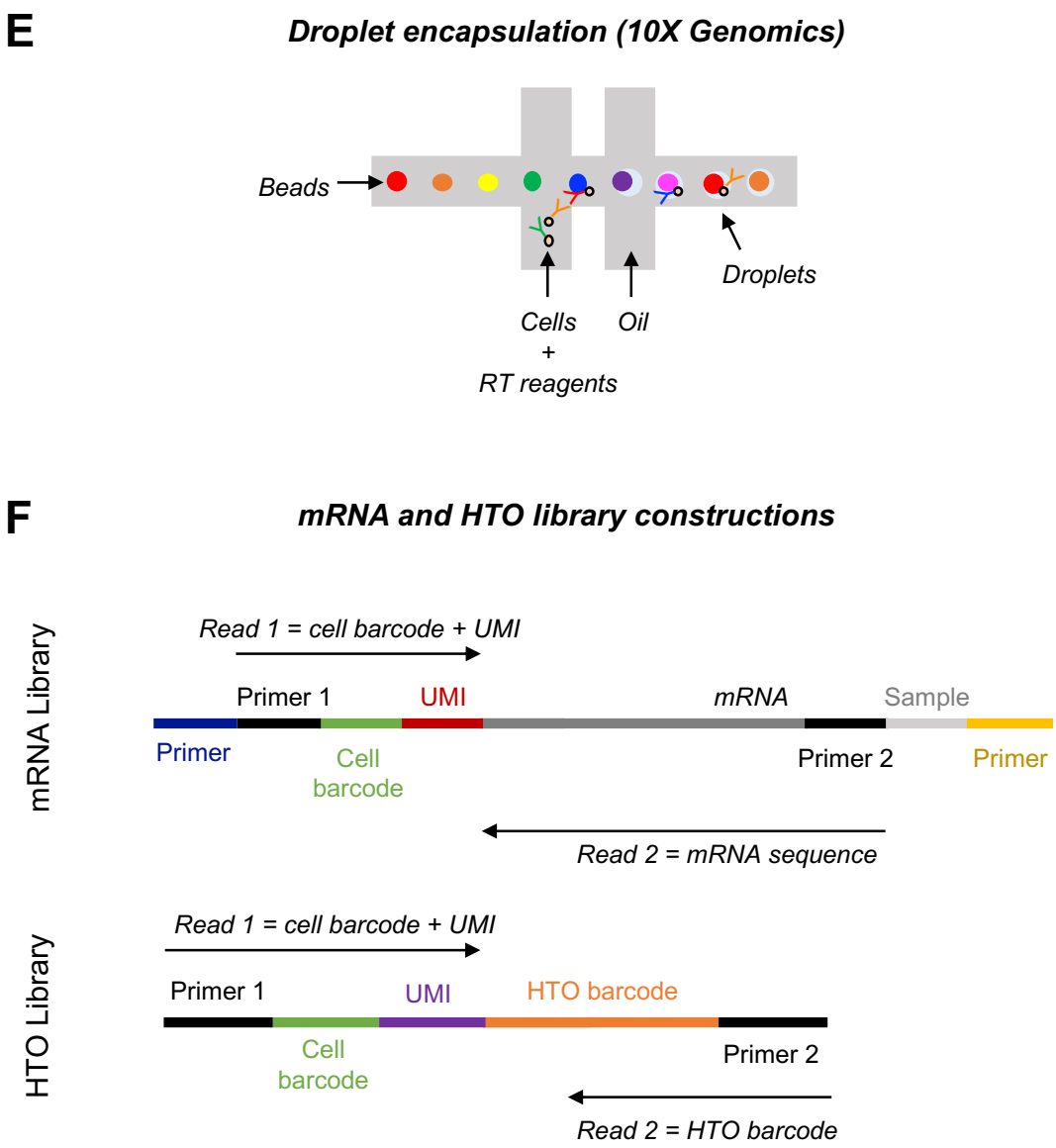


Figure 5.01. Overview of scRNA-Seq experiment

Continued from the previous page. (E) Using the 10X Genomics scRNA-Seq platform, individual HTO-tagged cells, reverse transcription (RT) reagents, and gel beads were encapsulated in oil – one cell per one bead. (F) Two sequencing libraries were constructed, one for the HTO/sample information, and one for the mRNA/transcript information. Each molecule in the mRNA library is read for information on 1) the identity of the gene (mRNA sequence), 2) the identity of the original transcript being read (unique molecular identifier or UMI), and 3) the identity of the cell the transcript was taken from. Each molecule in the HTO library is read for information on 1) the identity of the HTO (i.e. the sample the read derived from), 2) the identity of the cell the read derived from, and 3) a unique molecular identifier. By matching the cell barcode a given transcript is tagged with to the HTO that same cell barcode is tagged with, a cell can be assigned back to its original sample (WT_1, WT_2, etc.).

5.3 Pre-processing of the scRNA-Seq data[7]

Data analysis was conducted using the R package, 'Seurat' (version 3), according to the authors' standard workflow (282, 308). HTO expression data were normalized by a centered log ratio (computed independently for each gene) with the *NormalizeData* function, and then centered/scaled with the *ScaleData* function (282, 308). Following this, cells were demultiplexed by HTO enrichment to assign single cells back to their original sample (*HTODemux* function). 12,906 single cells, 3,883 doublets, and 1,515 cells negative for any HTO were recovered. Doublets and HTO-negative cells were omitted from further analysis (**Figure 5.02**). This left 3,853 WT_1 cells, 3,297 WT_2 cells, 2,763 KO_1 cells, and 2,993 KO_2 cells.

Following sample assignment, poor quality cells were removed from the dataset by filtering against the following criteria: 1) cells with fewer than 500 genes detected, 2) cells with over 30,000 RNA molecules detected (**unique molecular identifiers**, or **UMIs**), or 3) cells with over 6% of their total reads mapping to mitochondrial genes (**Figure 5.03**). This filtering aims to prevent empty droplets (i.e. those without cells), multiplets, and damaged cells from being incorporated into analyses. After these steps 3,853 WT_1 cells, 3,257 WT_2 cells, 2,721 KO_1 cells, and 2,949 KO_2 cells remained in the experiment (126 cells were removed).

Data from the 12,780 remaining cells were normalized and scaled with the *SCTransform* function, which transforms the data via a regularized negative binomial model. This function is also used to identify variable genes in the dataset (i.e. highly expressed in some cells, low in others). Variable genes are of the greatest biological interest for gene expressions studies; e.g. they could represent cell-type specific expression or the consequences of a gene KO. Lastly, unwanted sources of variation were regressed out (i.e. % mitochondrial reads and sequencing depth). Processed

[7] All data processing steps were performed by Moritz Haneklaus from the Livesey lab. Code for downstream analyses was also written (and kindly shared) by Moritz.

data then underwent linear dimensional reduction using the *RunPCA* command, and were clustered with *FindNeighbours* and *FindClusters*. Clustering was visualized by Uniform Manifold Approximation and Projection (UMAP), a non-linear dimension reduction technique (309).

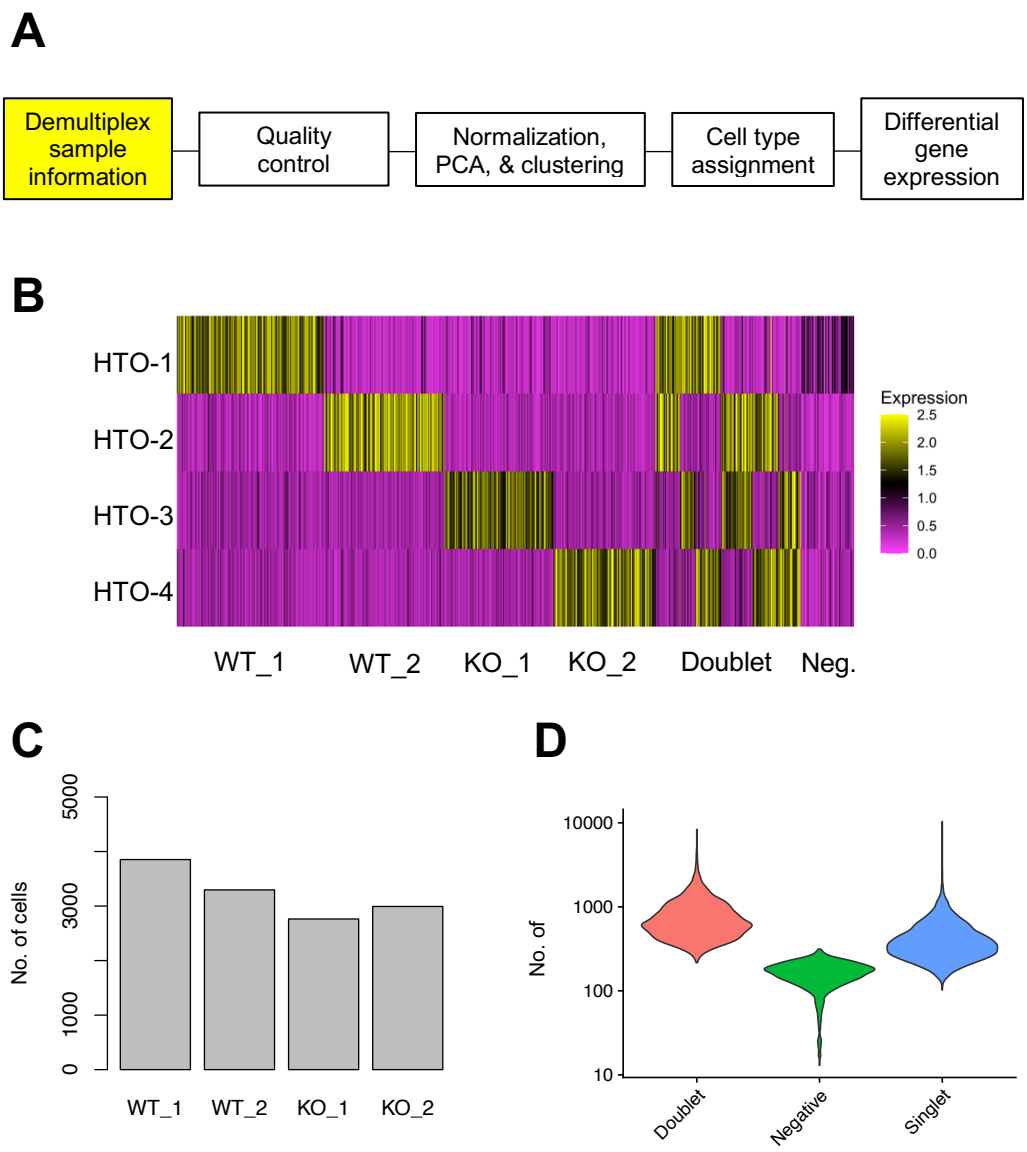


Figure 5.02. De-multiplexing of hashtag oligonucleotide-tagged samples

(A) Overview of the scRNA-Seq analysis pipeline. Yellow shading indicates the step in the pipeline that the current figure corresponds to. (B) Heatmap of scaled and normalized hashtag oligonucleotide (HTO) expression. WT induction 1 was tagged with HTO-1, WT induction 2 with HTO-2, KO induction 1 with HTO-3, and KO induction 2 with HTO-4. Based on the expression of an HTO, cells can be assigned back to their original sample. Cells that expressed more than one HTO are presumed doublets, while cells that expressed none of the HTOs were likely to be either empty droplets or cells that were poorly stained. 12,906 single cells, 3,883 doublets, and 1,515 HTO-negative cells were identified. Doublets and HTO-negative cells were omitted from downstream analyses. (C) After sample de-multiplexing 3,853 WT_1 cells, 3,297 WT_2 cells, 2,763 KO_1 cells, and 2,993 KO_2 cells were assigned. (D) Violin plot (log-scale) of the number of RNA molecules (UMIs) detected in cells belonging to doublet, HTO-negative, and singlet categories. As expected, doublets contained more RNA than singlets, while HTO-negative cells contained far fewer.

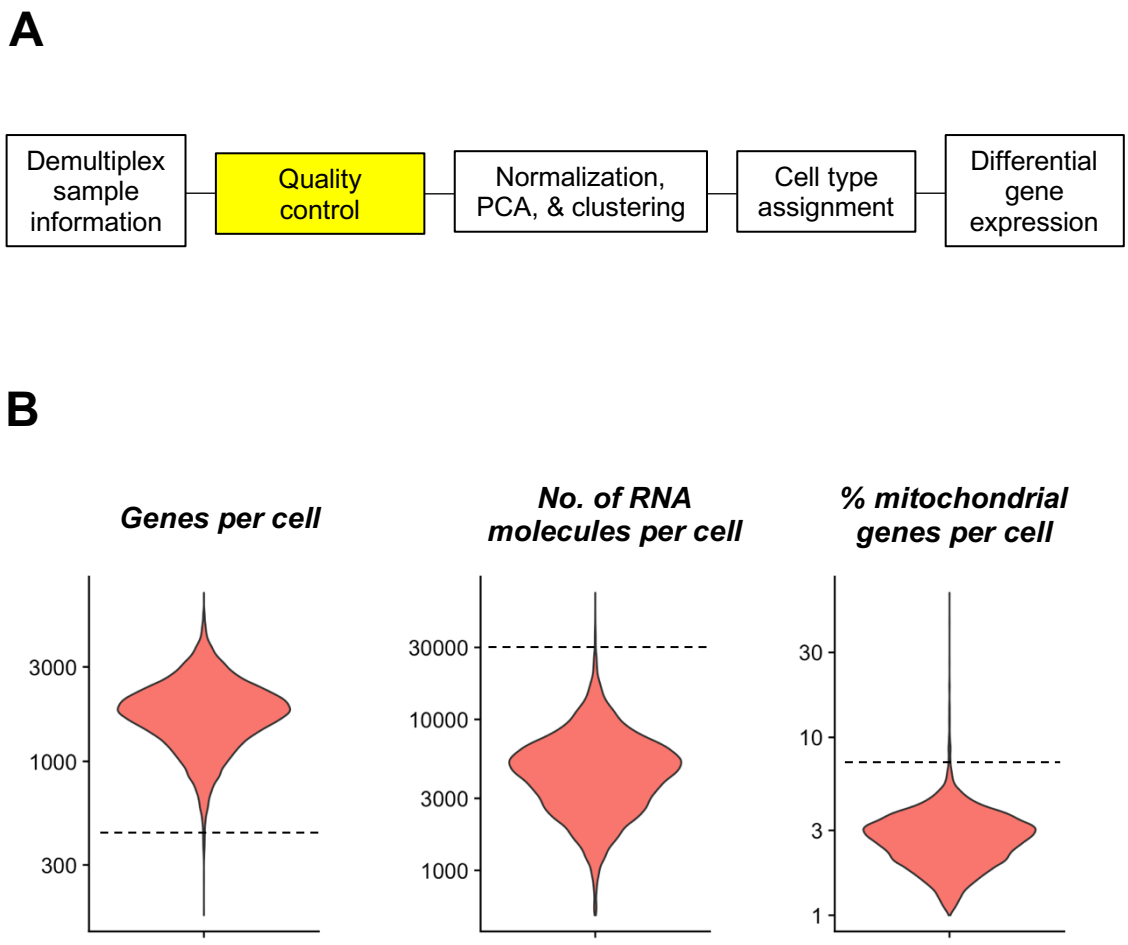


Figure 5.03. Pre-processing of scRNA-Seq data

(A) Overview of the scRNA-Seq analysis pipeline. Yellow shading indicates the step in the pipeline that the current figure corresponds to. (B) Following sample assignment, poor quality cells were removed from the dataset by filtering against the following criteria: 1) cells with fewer than 500 genes detected, 2) cells with over 30,000 RNA molecules detected (unique molecular identifiers, or UMIs), or 3) cells with over 6% of their total reads mapping to mitochondrial genes. This filtering aims to prevent empty droplets (i.e. those without cells), multiplets, and damaged cells from being incorporated into analyses

5.4 Assigning cell-type identifications to single cell clusters

Plotting the data with UMAP reduction revealed 9 distinct cell clusters (**Figure 5.04**). Visualizing the distribution of WT and KO samples amongst the clusters confirmed there were no obvious batch effects or significant loss of cell types between the genotypes (**Figure 5.05**). Genes that distinguished each cluster (i.e. cluster markers) were identified with the *FindMarkers* command (308), and used to assign cell type identities. Complete lists of marker genes for each of the 9 clusters can be found in **Appendix**.

*5.4.1 Cluster 1 – cortical progenitor cells (**Figure 5.06**)*

I first sought to categorize the clusters by cycling cells (e.g. progenitors) and post-mitotic cells (e.g. neurons). The *CellCycleScoring* function (308) revealed clusters 1, 8, and 9 were highly expressing genes associated with S, G2, and M phases (**Figure 5.05b**). As cluster 1 had a significant number of dividing cells, I first focused on identifying this population. Cells in this cluster robustly expressed forebrain/cortical markers like *FOXG1* (**Figure 5.07**), *EMX1, EMX2*, and *LHX2* (**Figure 5.09**) (187). Further investigation revealed these cells were strongly expressing progenitor-associated genes including *VIM*, *PAX6* (**Figure 5.10**), *SOX2, HES1,* and *NES* (173, 181, 310). A small number of cells in the group (< 20 cells) expressed the intermediate progenitor marker, *EOMES*. However, given the vast majority of cells were *EOMES*-negative, most of these progenitors are likely to be radial glia.

Sub-clustering (i.e. clustering only the cells in cluster 1) revealed three distinct populations within the group (**Figure 5.06**). These populations overlapped with cell cycle-associated genes, and not (as hypothesized) by maturity or *EOMES* expression. Moreover, existing scRNA-Seq publications have repeatedly shown that the cell cycle

is a strong driver of the clustering process, as this represents a major source of transcriptional variation across proliferative cells (310). As no other expression differences were found that could explain the division of these progenitors into distinct clusters, the following assignments were made: sub-cluster 0 = **G1/S cortical progenitors**; sub-cluster 1 = **S/G2 cortical progenitors**; and sub-cluster 2 = **G2/M cortical progenitors**.

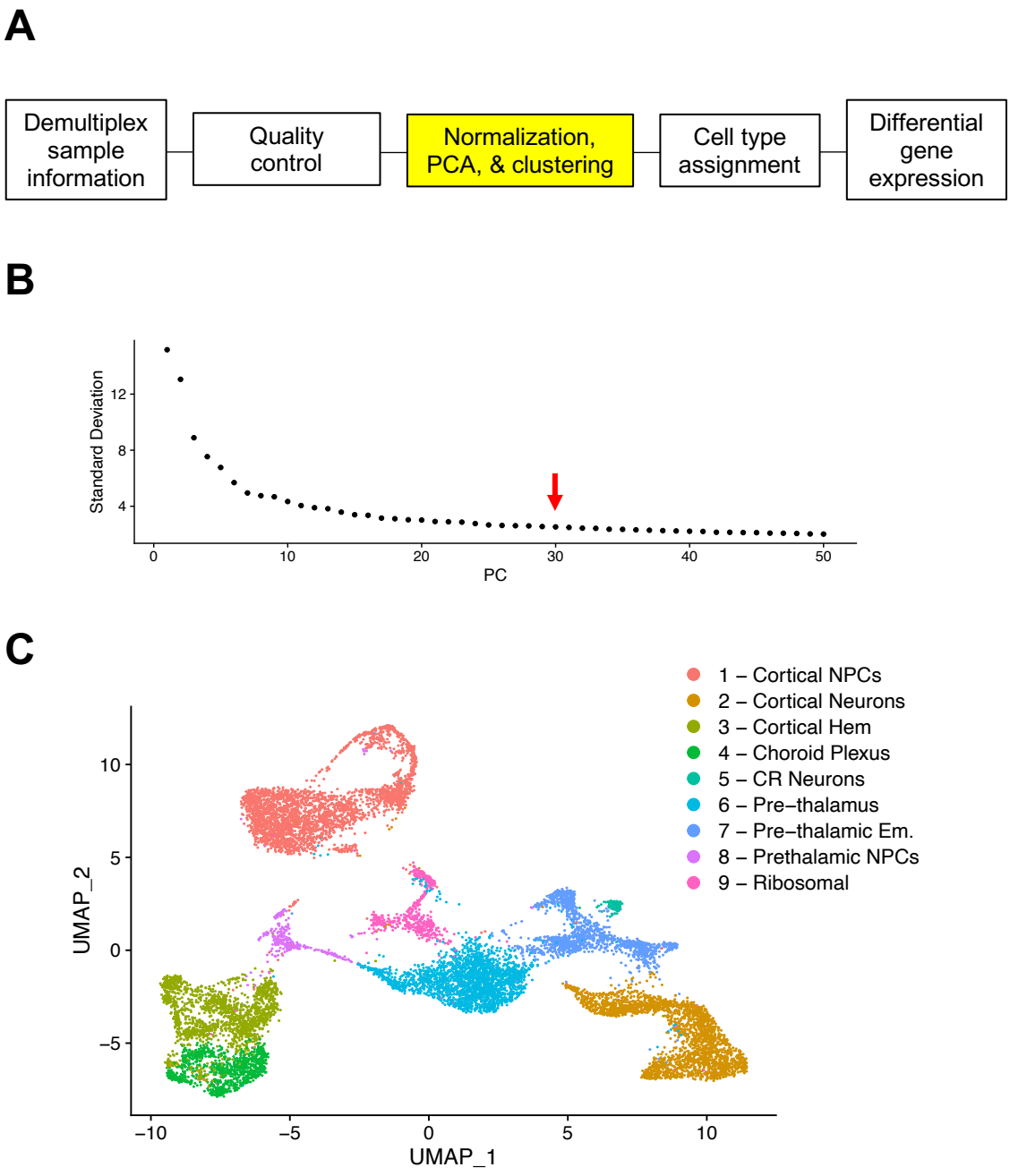


Figure 5.04. PCA and clustering of the dataset

(A) Overview of the scRNA-Seq analysis pipeline. Yellow shading indicates the step in the pipeline that the current figure corresponds to. (B) Elbow plot used to determine the 'dimensionality' of the dataset. The plot ranks principle components based on the percentage of variance each PC explains. An 'elbow' exists around PC10-30, suggests that the majority of true signal is captured in the first ~20-30 PCs. (C) Data from the 12,780 cells remaining after quality control, were clustered and visualized by Uniform Manifold Approximation and Projection (UMAP), a non-linear dimension reduction technique. 9 distinct clusters were identified.

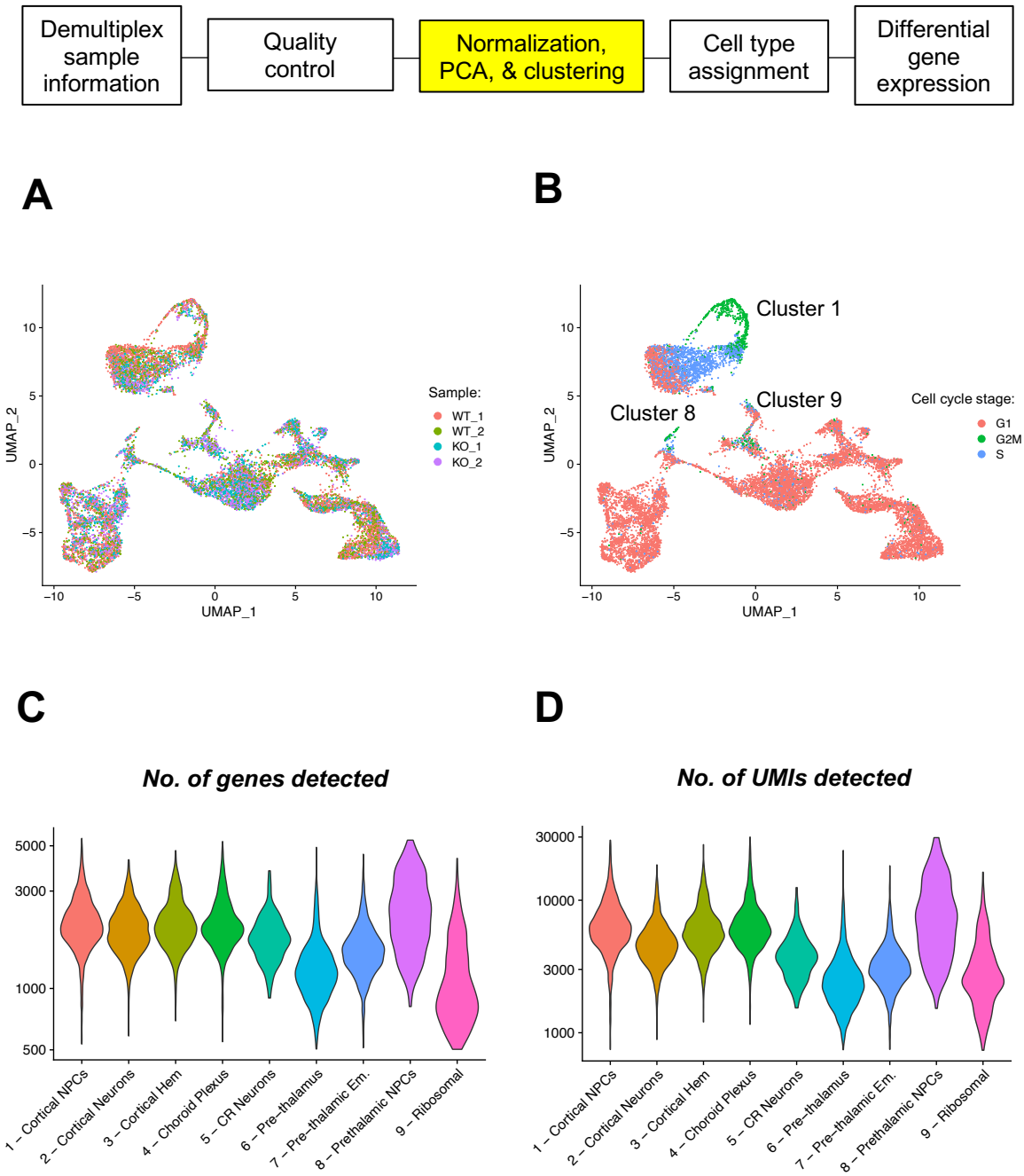


Figure 5.05. Clustering of the scRNA-Seq data

(A) Cells from all four samples appear in each of the clusters, confirming there were no significant batch effects or major differences in culture composition between genotypes. For further comparison of the cell types in WT versus KO samples please see **Figure 5.15**. (B) Each cell can be assigned to a cell cycle stage based on the expression of stage-associated genes. From this plot we can see clusters 1, 8, and 9 contain cells that are still cycling (i.e. progenitors). (C) For each cluster, the number of genes detected per cell and (D) the number of transcripts detected (UMIs) per cell are plotted. Clusters 6 and 9 have fewer genes and UMIs than the other clusters, while cluster 8 has modestly more.

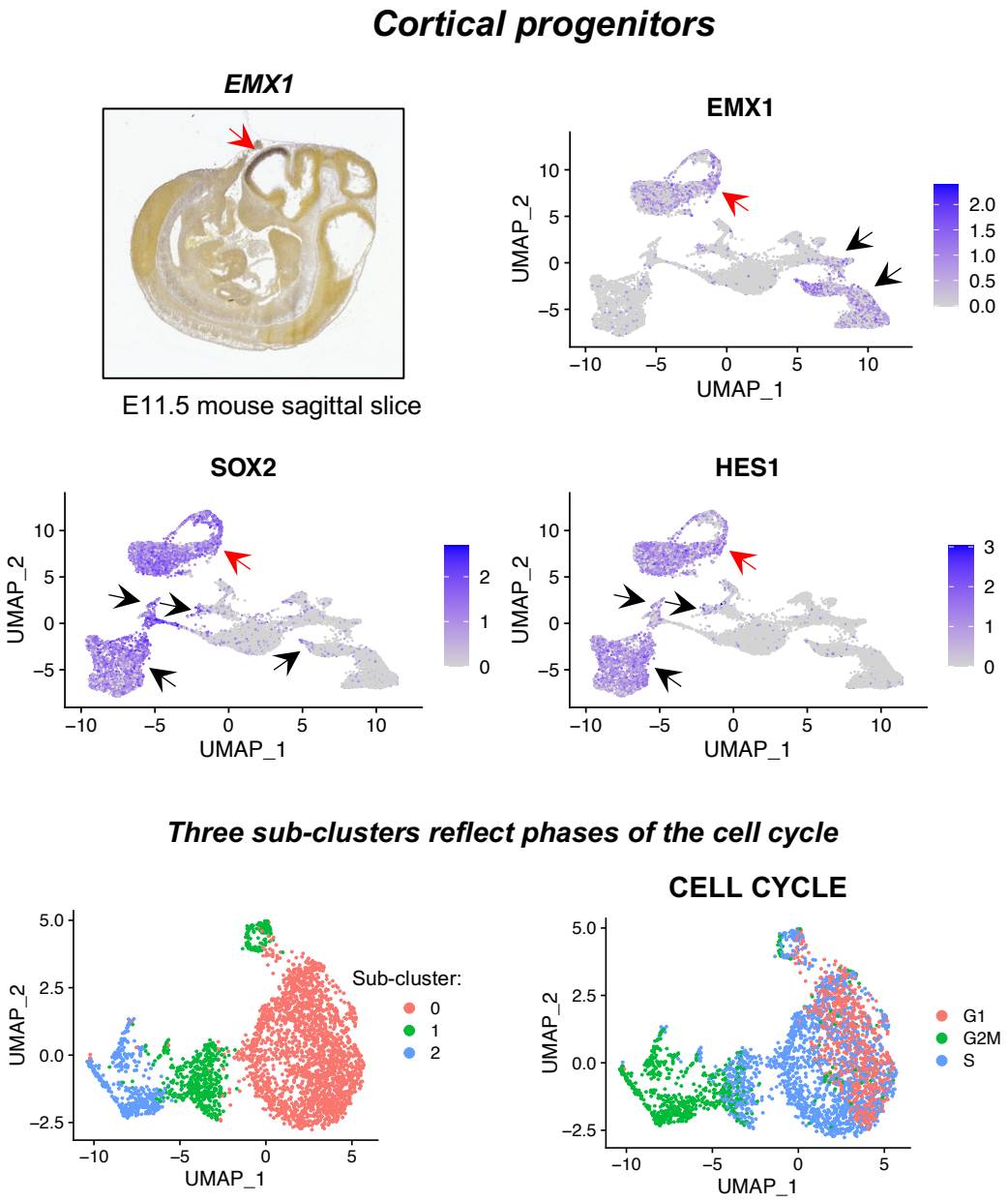


Figure 5.06. Cluster 1 – cortical progenitor cells (radial glia)

UMAP graphs of example genes distinguishing cluster 1 from the other eight clusters. In each graph, cluster 1 is highlighted by a red arrow. Black arrows show additional clusters that express the plotted gene. The cells in cluster 1 expressed established cortical progenitor markers, including *EMX1, SOX2,* and *HES1.* Further examination revealed there were three distinct sub-clusters within cluster 1. These cells grouped on the basis of the cell cycle, with sub-cluster 0 composed of cells at G1/S, sub-cluster 1 with cells at S/G2, and sub-cluster 2 with those at G2/M. A mouse sagittal section is shown for reference (Allen Developing Mouse Reference Atlas (Sunkin et al. 2013)). *In situ* hybridization demonstrates *EMX1* is specifically expressed in the cortex, confirming these progenitors are of cortical identity.

*5.4.2 Cluster 2 – deep layer cortical neurons (**Figure 5.07**)*

In addition to the cortical progenitor population, cluster 2 also expressed markers of the dorsal pallium/cortex (*EMX1:* **Figure 5.06** and *LHX2:* **Figure 5.09**) and telencephalon (*FOXG1*). Further inspection revealed cluster 2 was likely composed of **deep layer cortical neurons**. Specifically, these cells expressed markers of cortical layers 5-6 including *TBR1* (**Figure 5.11**), *CTIP2/BCL11B, SOX5*, and *FEZF2*. Crucially, they also expressed glutamatergic genes like *SLC17A7/VGLUT1* and *SLC17A6/VGLUT2*, further suggesting they were excitatory cortical neurons (187).

While all cells of cluster 2 expressed post-mitotic neuronal genes like *MAPT*, *MAP2 RBFOX3*, and *DCX* (310), there was a clear gradient in expression strength within the cluster. Sub-clustering identified three cell populations of increasing maturity. Sub-cluster 0 was the most mature, robustly expressing neuronal genes, while sub-clusters 1 and 2 showed weaker expression. Sub-cluster 1 was more mature than sub-cluster 2, given the latter was found to express the intermediate progenitor gene, *EOMES* , as well as pan-progenitor genes (e.g. SOX2, **Figure 5.06**). As such, the cells in sub-cluster 2 are likely to be at the final stages of transition between the intermediate progenitor and post-mitotic neuronal fates. The decision to call them neurons was underpinned by 1) their expression of neuronal genes, 2) their clustering within a larger neuronal population, and 3) the fact that they were not cycling (i.e. were in G1 phase).

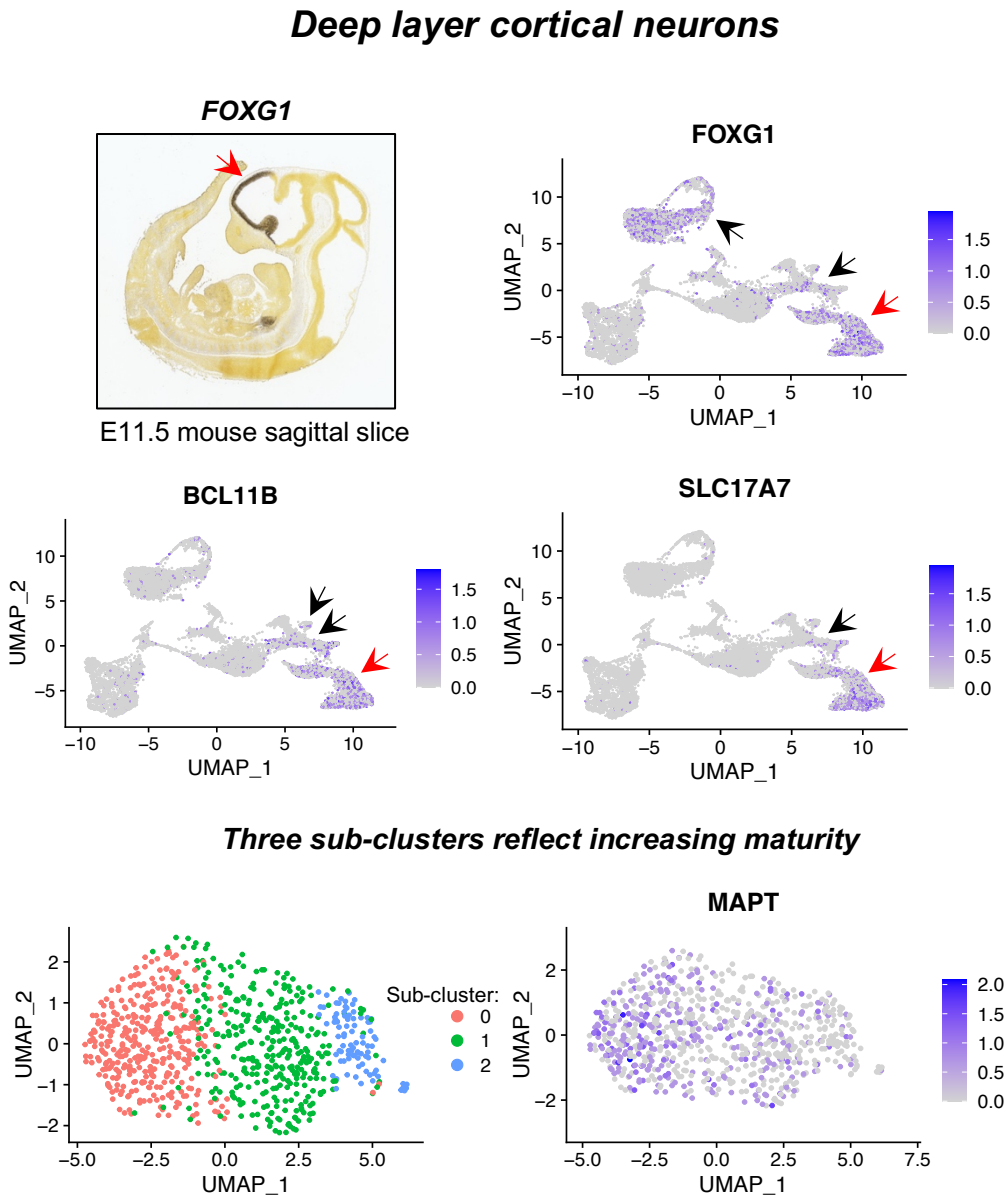


Figure 5.07. Cluster 2 – deep layer cortical neurons

UMAP graphs of example genes distinguishing cluster 2 from the other eight clusters. In each graph, cluster 2 is highlighted by a red arrow. Black arrows show additional clusters that express the plotted gene. The cells in cluster 2 expressed established cortical neuron markers, including *BCL11B (CTIP2),* and *SLC17A7 (VGLUT1).* Further examination revealed there were three distinct sub-clusters within cluster 2. These cells grouped on the basis of maturity, with sub-cluster 0 composed of the most mature neurons (highest expression of neuronal marker, *MAPT*), sub-cluster 1 less mature, and sub-cluster 2 least mature. A mouse sagittal section is shown for reference. *In situ* hybridization demonstrates how *FOXG1* is specifically expressed in the telencephalon, confirming these neurons are of telencephalic identity.

*5.4.3 Clusters 3 and 4 – progenitors of the cortical hem and choroid plexus (**Figure 5.08**)*

Clusters 3 and 4 expressed *RSPO2*, *LMX1A*, and *WNT3A* – all of which are expressed predominantly by the cortical hem. Cluster 4 further expressed *TTR,* a marker of the choroid plexus and which neighbors the hem *in vivo*. Neither the hem nor choroid plexus express *FOXG1*, despite their location within the telencephalon (311, 312). This would therefore explain why clusters 3 and 4 are *FOXG1*-negative. Furthermore, both clusters expressed pan-progenitor markers *(*see *SOX2*, **Figure 5.06***)* but did not robustly express neuronal markers (*DCX/MAPT*). As such, the majority of cells in these groups are likely progenitor cells and were assigned the following identities: **cortical hem progenitors** (cluster 3) and **choroid plexus progenitors** (cluster 4). No meaningful sub-clusters were identified in either the hem or choroid plexus populations.

*5.4.4 Cluster 5 – Cajal-Retzius (CR) neurons (**Figure 5.09**)*

Cluster 5 was found to express *TP73*, a marker of **Cajal-Retzius (CR) neurons** (313). Cells in this cluster also expressed *RELN*, *CALB2,* and *EBF3* – three other markers of CR neurons (314). Moreover, Cajal-Retzius cells have been shown to derive from different regions in the developing telencephalon. These include the cortical hem, choroid plexus, septum, and anti-hem. For a few key reasons, cluster 5 is most likely composed of cells from the cortical hem or choroid plexus. Firstly, CR neurons derived from the anti-hem or septum do not express *TP73*. These CR cells do, however, express *DBX1/GSX2* (anti-hem derived) or *ETV1* (septum derived), both of which were not expressed in cluster 5 (315, 316). Moreover, cluster 5 robustly expressed markers of the dorsal pallium (*LHX2*, etc.), where CR neurons migrate to during their development. Lastly, expression of pan-neuronal markers *(DCX/MAPT)* and glutamatergic markers (*SLC17A6*) confirmed these cells were glutamatergic neurons as are Cajal-Retzius cells (317).

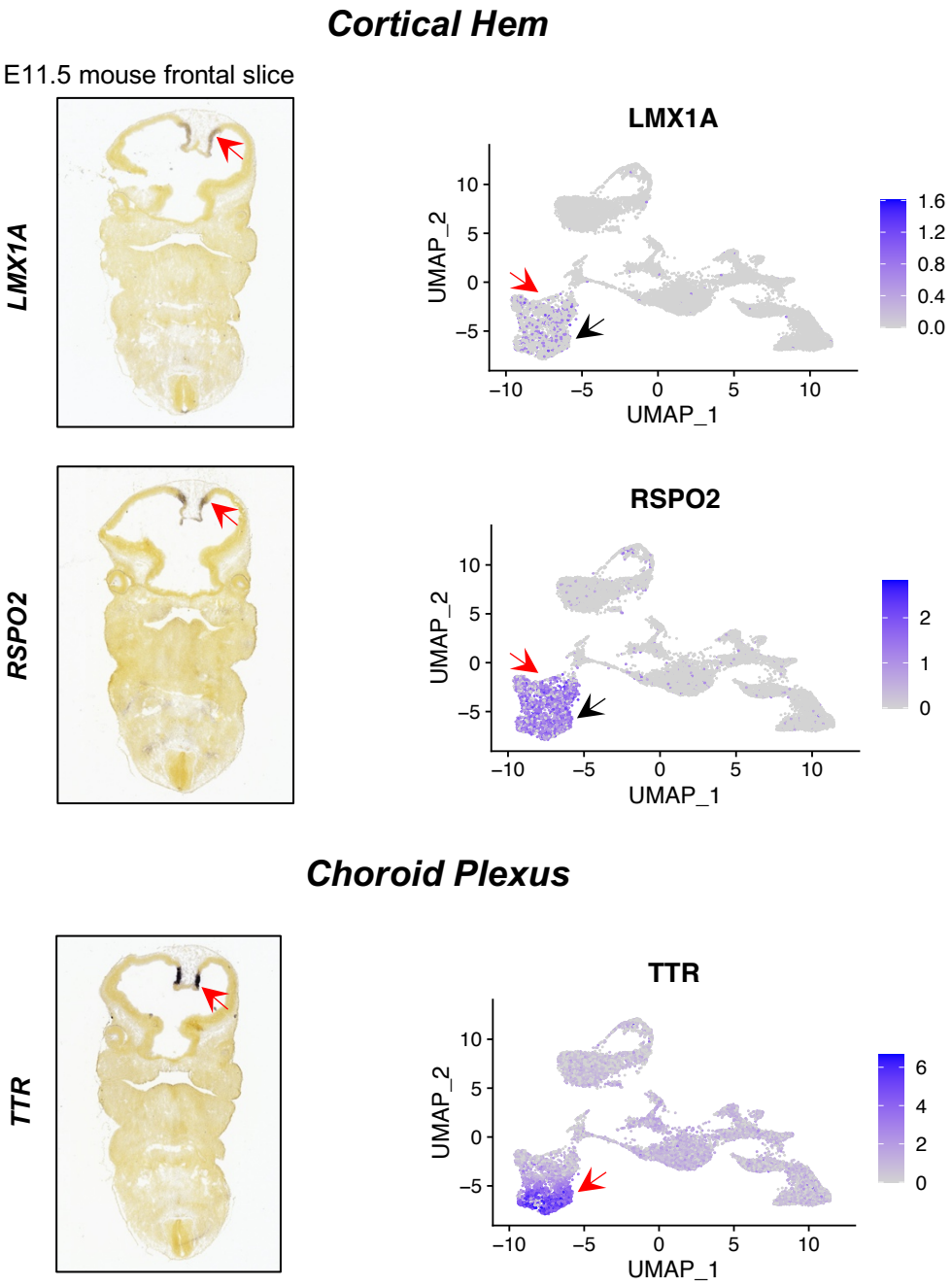


Figure 5.08. Clusters 3 (cortical hem) and 4 (choroid plexus)

UMAP graphs of example genes distinguishing clusters 3 and 4 from the other seven clusters. In each graph, cluster 3 or 4 is highlighted by a red arrow. Black arrows show additional clusters that express the plotted gene. The cells in cluster 3 expressed established cortical hem markers, including *LMX1A* and *RSPO2*, while cluster 4 expressed the choroid plexus marker, *TTR*. A mouse frontal section is shown for reference, with *in situ* hybridizations showing the restricted expression of *LMX1A, RSPO2,* and *TTR* to the hem/choroid plexus.

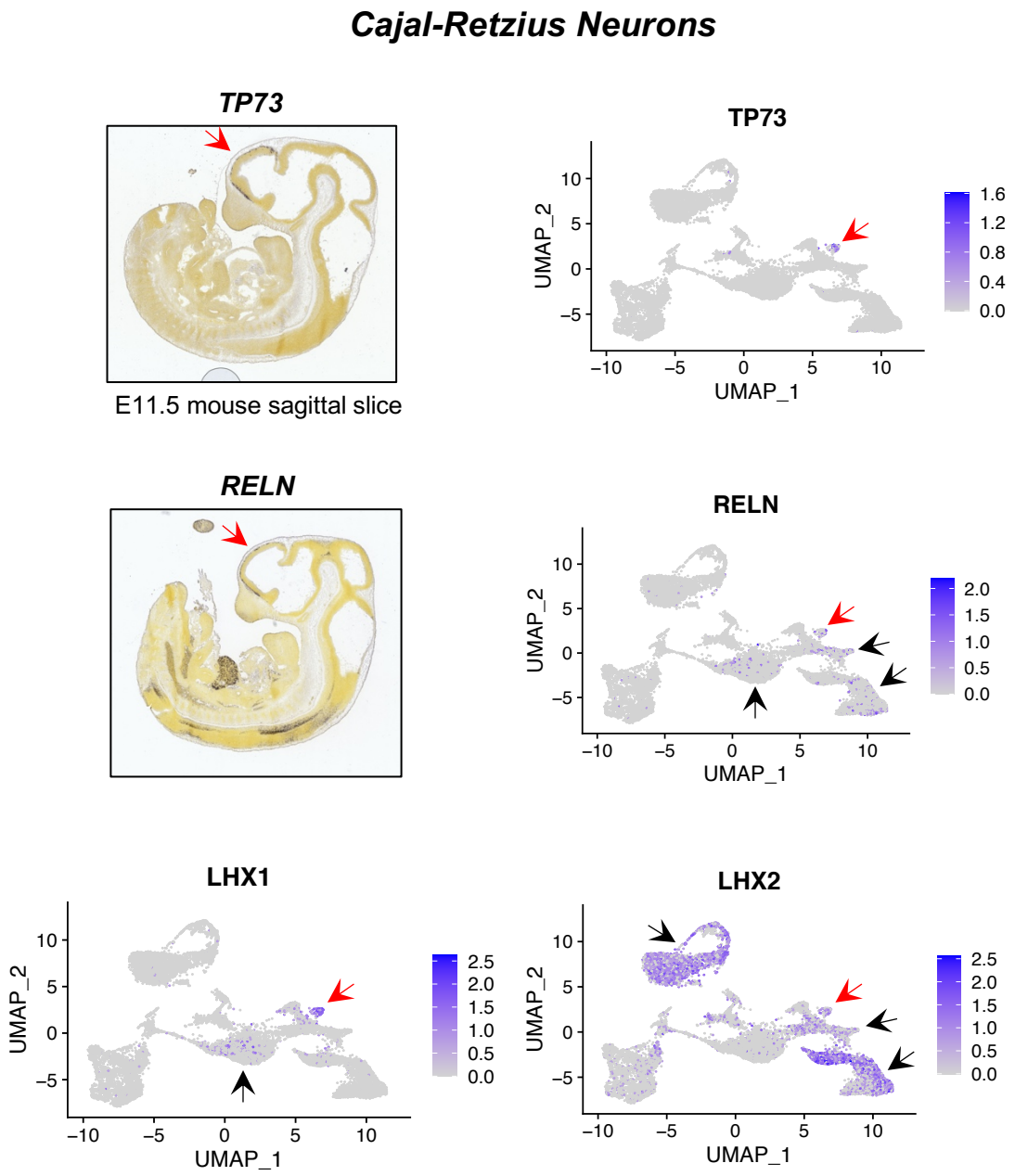


Figure 5.09. Cluster 5 – Cajal-Retzius (CR) neurons

UMAP graphs of example genes distinguishing cluster 5 from the other eight clusters. In each graph, cluster 5 is highlighted by a red arrow. Black arrows show additional clusters that express the plotted gene. The cells in cluster 5 expressed established Cajal-Retzius markers, including *TP73, RELN,* and *LHX1.* Crucially, these cells also expressed cortical markers *(LHX2),* confirming they were cortical CR neurons. A mouse sagittal section is shown for reference, with *in situ* hybridizations showing *RELN* and *TP73* expression in the cortex (and telencephalon more widely).

*5.4.5 Cluster 6 – Pre-thalamus (**Figure 5.10**)*

Cluster 6 expressed classical interneuron markers like *GAD1/2*, *DLX1/2/5/6*, and *SLC32A1 (VGAT)* (187). However, it did not express *FOXG1* (**Figure 5.07**), ruling out the possibility these cells derived from the ganglionic eminence (the source of cortical interneurons). Cluster 6 also expressed diencephalic genes like *SIX3, ISL1,* and *ARX*, reinforcing these cells were not of telencephalic identity. Other strong markers of this cluster included *PAX6, MEIS2*, *NR2F1* – which the Allen Developing Mouse Atlas (318) identified as overlapping in the developing pre-thalamus. This structure lies adjacent to the telencephalon *in vivo*, and therefore could quite plausibly be generated by our forebrain patterning protocol *in vitro* (see **Figure 5.14**). Furthermore, 'caudalized' differentiations have been previously generated in our lab (187).

*5.4.6 Cluster 7 – Caudal pallium / pre-thalamic eminence (**Figure 5.11**)*

Cluster 7 was composed of two distinct sub-populations. Sub-cluster 0 strongly expressed *EMX1* (**Figure 5.06**), *NEUROD2/6*, and *FOXG1* (**Figure 5.07**). Sub-cluster 1 expressed *LHX1*/5, *NR2F2*, and *SIX3*. The Allen Developing Mouse Atlas (318) revealed that two adjoining brain regions express these distinct groups of markers *in vivo*: the pallium and the pre-thalamic eminence (PThE) (see **Figure 5.14**). The PThE expresses *TBR1*, *EOMES*, *LHX9*, and *CALB2* (319)– which are all observed in sub-cluster 1. Similarly, the PThE is glutamatergic, and cluster 7 expresses *SLC17A6 (VLUT2)* and *GRIA* (319).

On the other hand, *EMX1* and *FOXG1* expression strongly suggests sub-cluster 0 contains cells from the dorsal pallium. It is possible cluster 7 has been 'mis-clustered' by the Seurat algorithm, and contains two distinct populations of cells. However, it is more likely that cluster 7 represents cells belonging to a boundary of pallium and PThE.

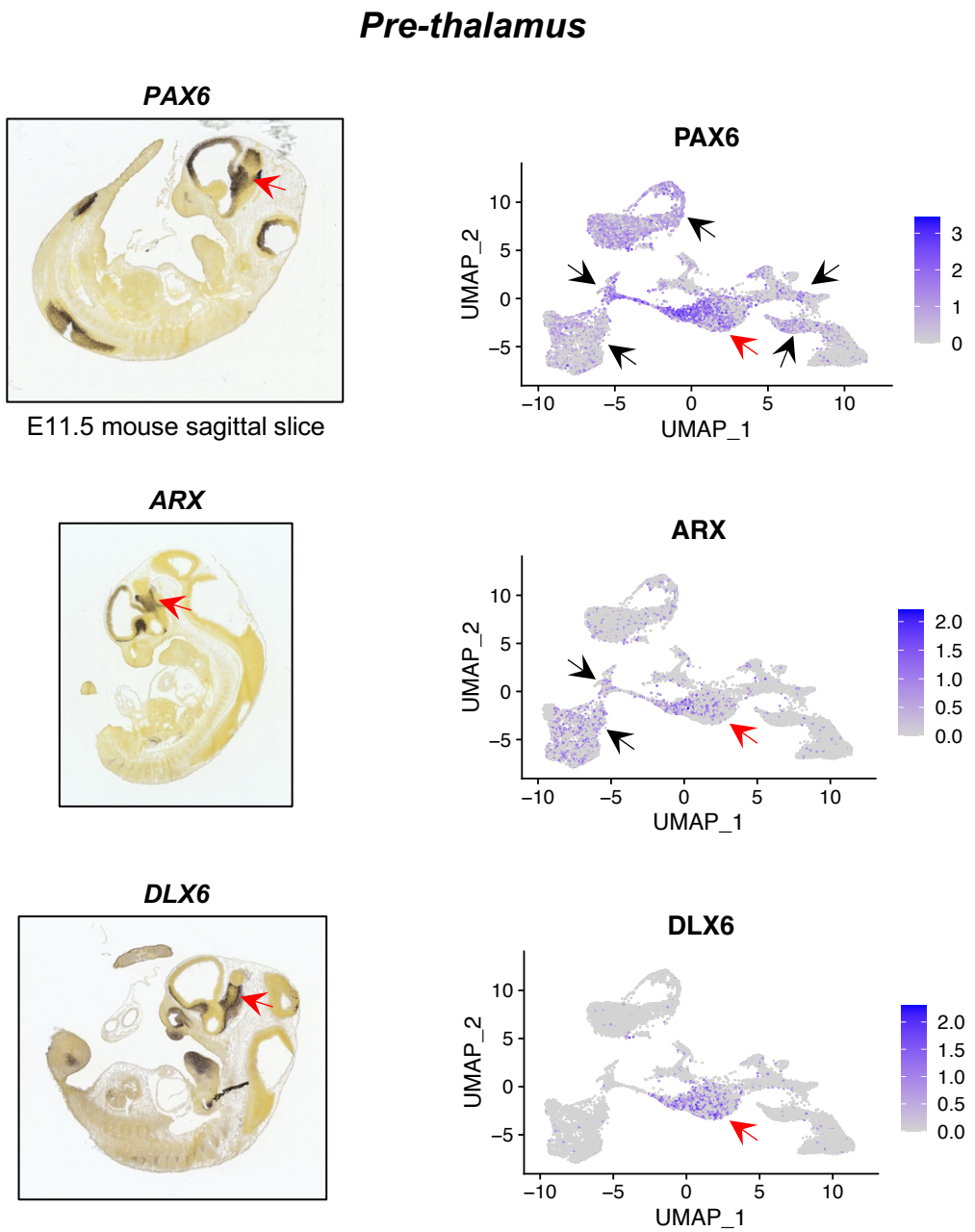


Figure 5.10. Cluster 6 – pre-thalamus (diencephalon)

UMAP graphs of example genes distinguishing cluster 6 from the other eight clusters. In each graph, cluster 6 is highlighted by a red arrow. Black arrows show additional clusters that express the plotted gene. The cells in cluster 6 expressed established pre-thalamic markers, including *PAX6, ARX,* and *DLX6.* Mouse sagittal sections show *in situ* hybridizations of these three markers. As evident, the only region of overlap between the markers is the pre-thalamus (red arrow). The lack of *FOXG1* (telencephalon marker) but expression of diencephalic genes like *ARX* and *SIX3,* further support this cell type assignment.

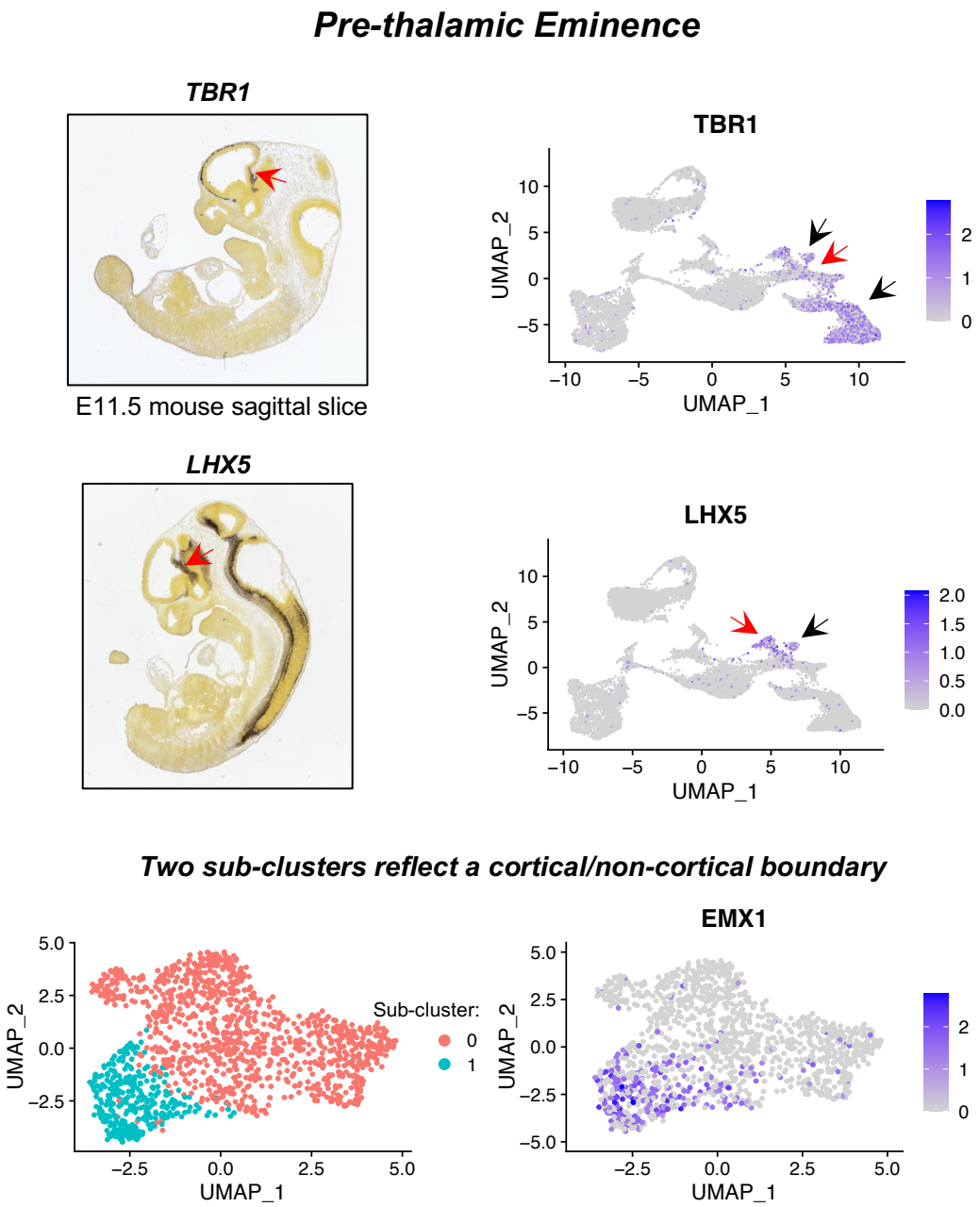


Figure 5.11. Cluster 7 – Prethalamic eminence (telencephalon/diencephalon border)

UMAP graphs of example genes distinguishing cluster 7 from the other eight clusters. In each graph, cluster 7 is highlighted by a red arrow. Black arrows show additional clusters that express the plotted gene. The cells in cluster 7 expressed markers of the pre-thalamic eminence (PThE), including *TBR1* and *LHX5.* Mouse sagittal sections show *in situ* hybridizations of these two markers. As evident, *TBR1* and *LHX5* expression overlaps in the pre-thalamic eminence (red arrow). Two distinct sub-clusters were identified in the population. Sub-cluster 0 expressed PThE markers, while cluster 1 also expressed cortical markers (e.g. *EMX1*). Given the grouping of these sub-clusters together, they could represent a gradient of the caudal cortex/PThE boundary.

*5.4.7 Cluster 8 – Pre-thalamic progenitors (**Figure 5.12**)*

Cluster 8 expressed *GSX2* and *SFRP2*, two genes commonly used as markers of the cortical anti-hem (320). However, cluster 8 did not express *FOXG1* (**Figure 5.07**) as with cluster 6. Another region that expresses the markers found in cluster 8 was the pre-thalamus, and particular progenitor cells. *GSX2*, *WNT7B*, *ARX* (**Figure 5.10**), *GDF10*, and *PAX6* are all markers of cluster 8 and are expressed by pre-thalamic progenitors (321, 322). Cluster 8 also expressed pan-progenitor genes like *VIM, SOX2* (**Figure 5.06**), and *FABP7*. The *CellCycleScoring* function showed cells in this group were in S and G2/M phases, which also suggests they are dividing progenitors (**Figure 5.05b**).

*5.4.8 Cluster 9 – Ribosomal gene-enriched cells (**Figure 5.13**)*

The *FindMarkers* function in Seurat returned markers for cluster 9 that were highly enriched for ribosomal genes. The few genes that were not ribosomal, were somehow associated with translation (e.g. pre-mRNA splicing, tRNA methyltransferase, elongation factors, and/or histone proteins). No other pattern could be deduced from these markers. Cell cycle analysis revealed cells were in G1, S, and G2/M stages, but only a small proportion of the cluster expressed progenitor genes (*SOX2* and *HES1*: **Figure 5.06**). In fact, cluster 9 showed limited expression of most genes plotted in the UMAP graphs, regardless of cell type or maturity. This observation can also be noted from **Figure 5.06c-d**, where cluster 9 clearly has a lower number of genes and UMIs detected than most of the other clusters. It is possible cluster 9 represents dying cells that do not exhibit an unusually high percentage of mitochondrial genes (but showing many ribosomal genes instead) (323). However, cluster 9 did not strongly express markers of apoptosis (e.g. p53) or DNA damage (e.g. JUN).

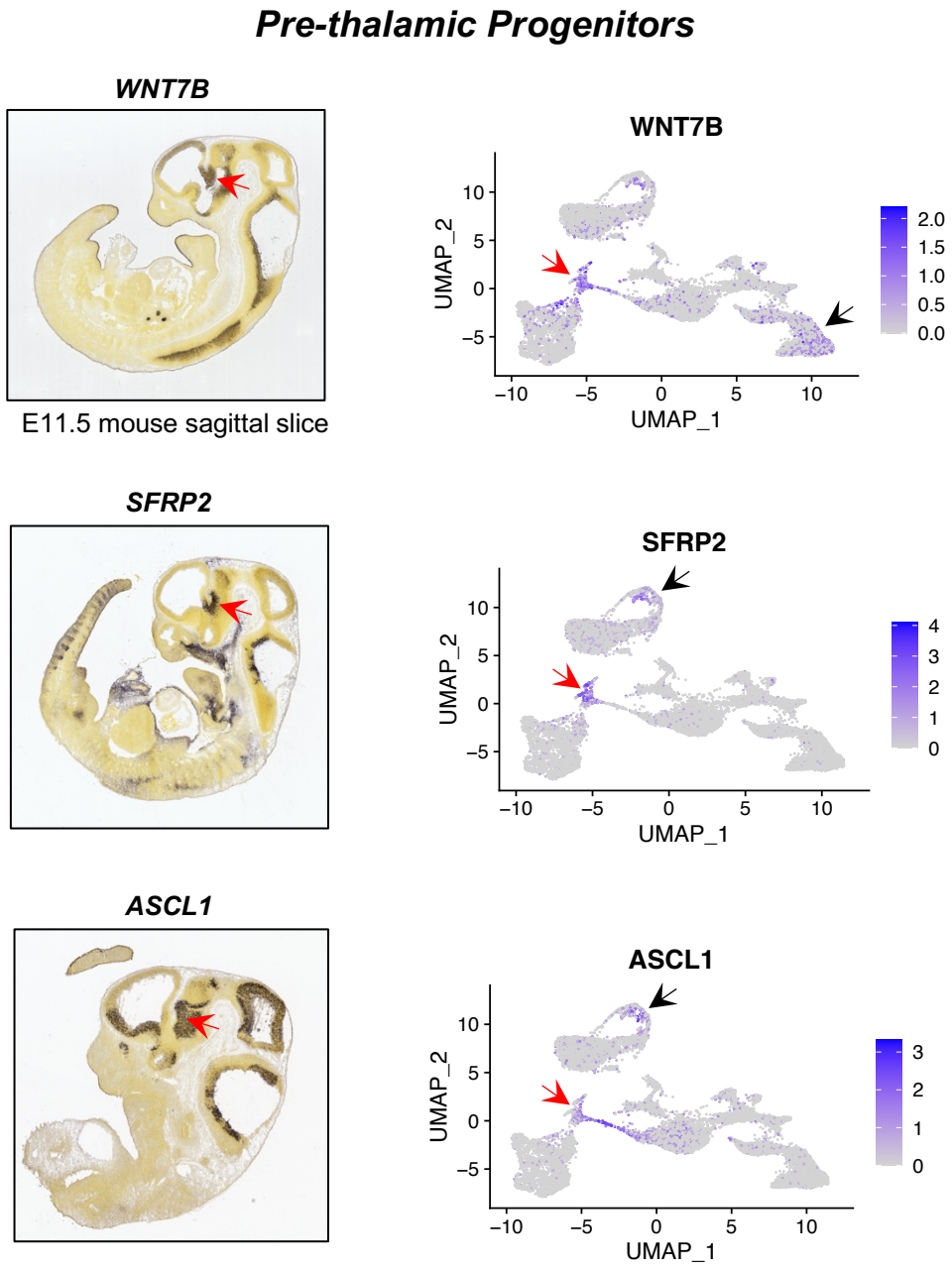


Figure 5.12. Cluster 8 – Pre-thalamic progenitors

UMAP graphs of example genes distinguishing cluster 8 from the other eight clusters. In each graph, cluster 8 is highlighted by a red arrow. Black arrows show additional clusters that express the plotted gene. The cells in cluster 8 expressed pre-thalamic progenitor markers, including *WNT7B, SFRP2,* and *ASCL1.* Mouse sagittal sections show *in situ* hybridizations of these three markers. As evident, the only region of overlap between the markers is the pre-thalamus (red arrow). These cells expressed pan-progenitor genes (e.g. *SOX2, ASCL1*), and did not express pan-neuronal genes (e.g. *MAPT*), supporting their assignment as progenitor cells.

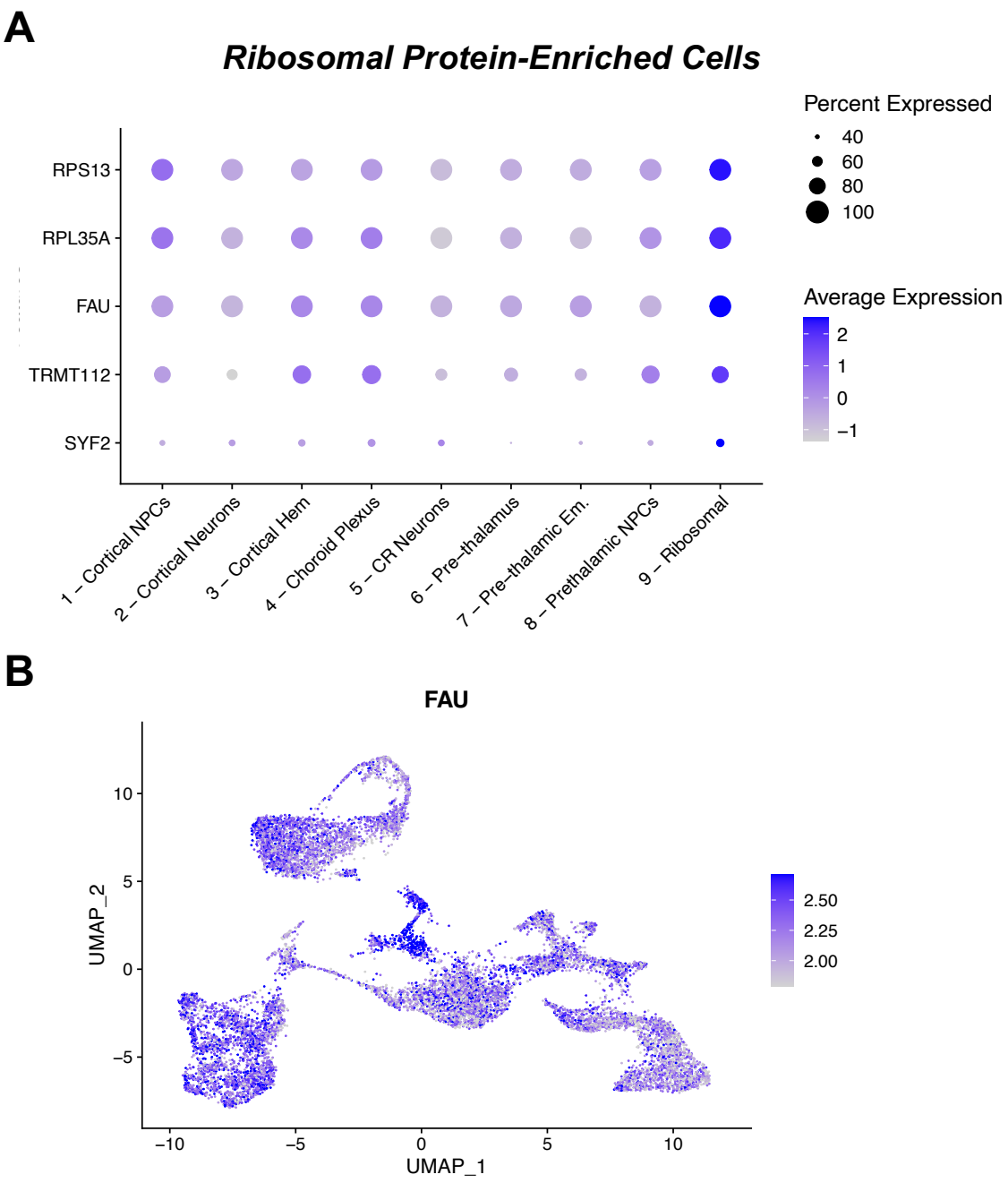


Figure 5.13. Cluster 9 – Ribosomal protein-enriched cells

(A) Dotplot of genes distinguishing cluster 9 from the other eight clusters. The cells in cluster 9 were highly enriched in genes encoding ribosomal proteins (e.g. *RPS13, RPL35A,* and FAU). Other enriched genes included proteins involved in translation. For example, *TRMT112* encodes a tRNA methyltransferase subunit and *SYF2* a pre-mRNA splicing factor. No other meaningful pattern in the cluster 9 marker genes were identified. (B) UMAP plot of the ribosomal protein, *FAU* (contrast adjusted). The strongest expression is noted in cluster 9.

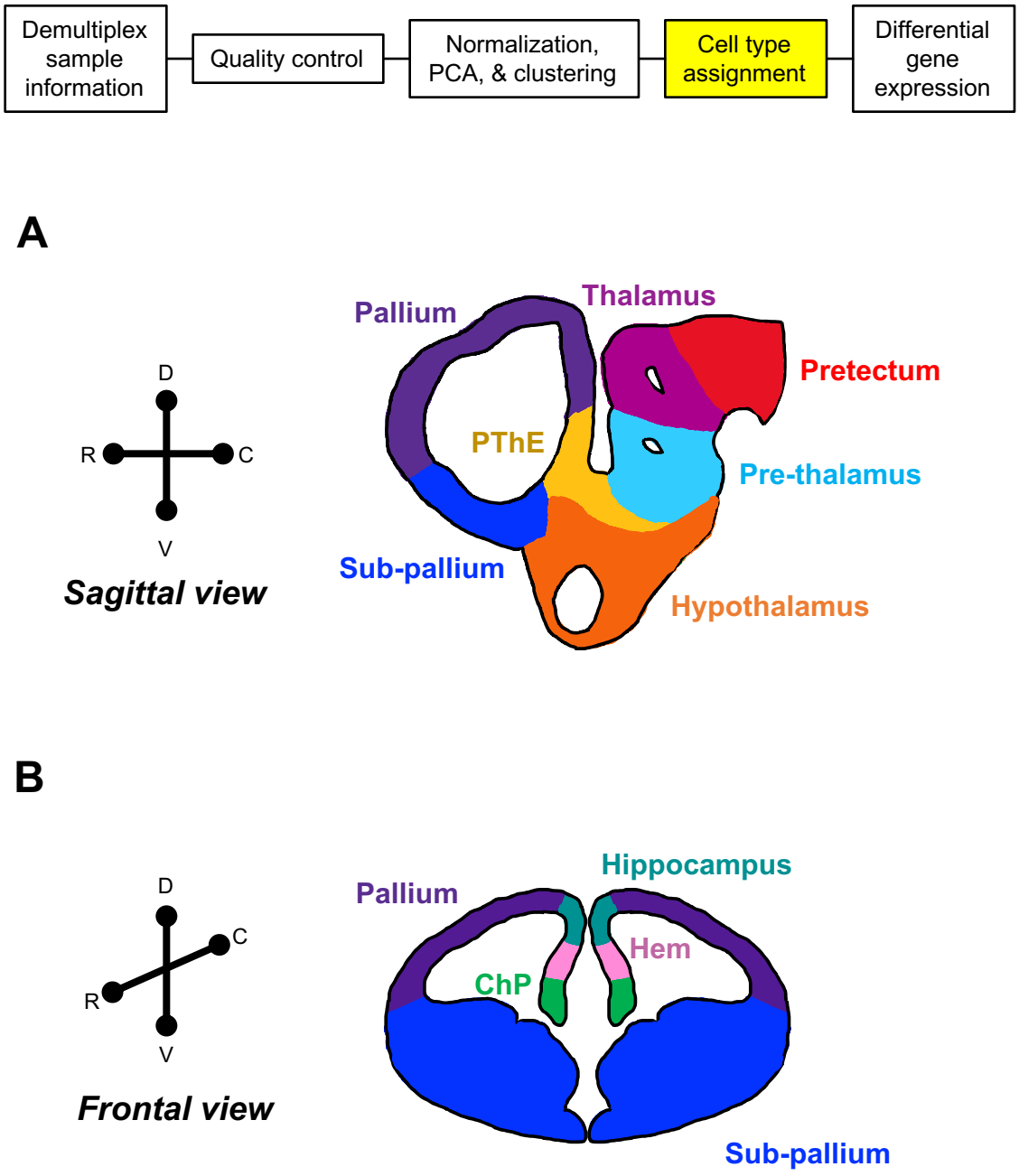


Figure 5.14. Schematic of an E11.5 mouse forebrain

(A) Sagittal section of the telencephalon and anterior diencephalon of an E11.5 mouse (figure adapted from the Allen Developing Mouse Reference Atlas (Sunkin et al. 2013)). The pallium contains cortical progenitors, cortical neurons, and Cajal-Retzius neurons. The pre-thalamic eminence (PThE) and pre-thalamus immediately neighbor the pallium, and are therefore plausible cell types to be generated by our cortical induction protocol. (B) Frontal section of an E11.5 mouse telencephalon. The cortical hem and choroid plexus (ChP) lie just ventromedially to the pallium. **R:** rostral; **D:** dorsal; **C:** caudal; **V:** ventral.

5.5 Cluster-level distribution of WT and KO cells

As shown in **Figure 5.05a**, there were no macro-level differences in culture composition between the *CNTNAP2* WT and KO lines (i.e. the complete loss of the gene does not lead to profound differences in cell types at D50). As expected for cell lines that were differentiated simultaneously, all clusters contained cells from both genotypes. Perhaps more surprising, was the fact that two independent inductions of paired WT and KO lines were both caudalized (i.e. producing diencephalic structures). However, emerging research has shown certain cell lines have an intrinsic tendency to produce ventralized and/or caudalized cortical inductions (187). That being said, I wanted to clarify whether there were differences in the proportion of WT and KO cells in each cluster. **Figure 5.15** shows the relative contributions of the four inductions to each of the nine clusters. In both WT and KO lines cortical progenitors were the most common cell type, representing 3014 cells or 24% of all cells in the dataset. Of these 3014 cells, 67% (2033 cells) were from WT samples and 33% (981 cells) from the KO samples. The next largest population was post-mitotic cortical neurons, which were 2410 or 18.9% of all cells. As would be expected, the ratio of WT to KO cells were highly similar to that of the cortical progenitor cluster, 65% and 35% respectively.

Pre-thalamic neurons were the third most common cell type (2347 cells/18.4% of the 12,753 cells), followed by cortical hem (1666 cells/13%), and pre-thalamic eminence (1372 cells/10.8%). All three clusters had a relatively similar distribution of the two genotypes, with 42% to 58% (pre-thalamus), 45% to 55% (hem), and 54% to 46% (pre-thalamic eminence) of WT to KO cells. There were 954 cells of the choroid plexus (7.5% of the dataset), with 64% WT and 35% KO. The least common clusters were pre-thalamic progenitors (304 cells/2.4%), Cajal-Retzius neurons (109 cells/0.9%), and the ribosomal gene-enriched cells (577 cells/4.5%). Lastly, within these clusters 64% of the choroid plexus was WT (36% KO), 46% of the pre-thalamic progenitors were WT (54% KO), and 39% of the ribosomal cluster were WT (61% KO). A Welch's *t*-test detected the WT to KO ratio as mildly significant in the

Cajal-Retzius cluster ($p = 0.007$). No other ratios reached significance, suggesting the overall composition of the WT and KO cultures were relatively similar.

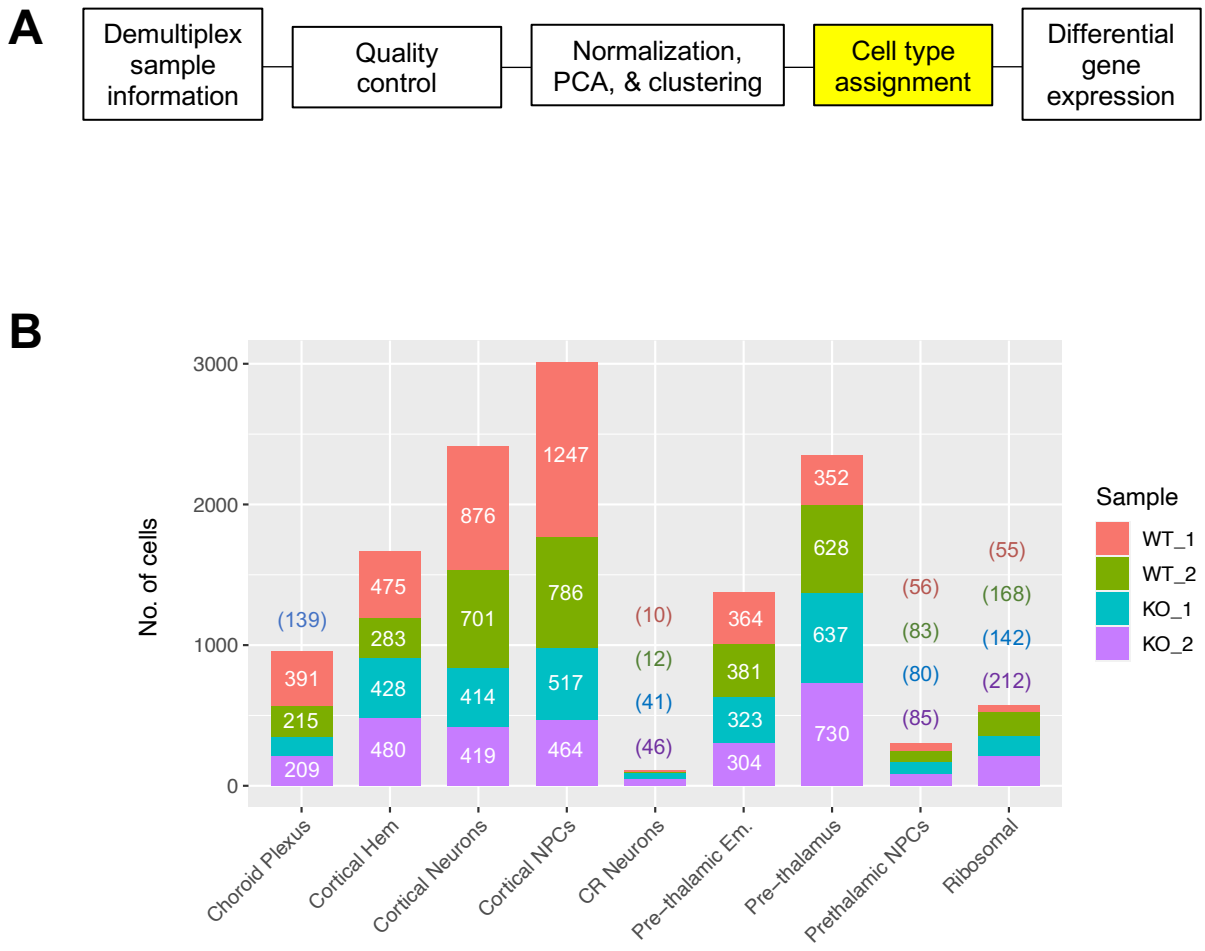


Figure 5.15. Cluster compositions by genotype

The number of cells in each cluster are plotted by sample (cell counts shown directly in the graph). In both WT and KO lines, cortical progenitors were the most common cell type at 3014 cells/24% of the dataset. The least common cell type was Cajal-Retzius neurons, comprising only 109 cells/0.9%. A Welch's *t*-test found the Cajal-Retzius neurons were composed of significantly more KO cells than WT cells ($p = 0.007$). No other significant differences in the proportions of WT to KO cells were identified in the remaining eight clusters.

5.6 *CNTNAP2 expression in PSC-derived cortical cultures at single cell resolution*

5.6.1 Cell type-specific CNTNAP2 expression

Figure 5.16 shows the expression of *CNTNAP2* in each of the nine cell types. For the purposes of analyzing normal *CNTNAP2* expression in our cortical cultures, only WT cells were considered. The highest expression of *CNTNAP2* is observed in deep layer cortical neurons (cluster 2). Specifically, the most mature cells in this cluster (sub-cluster 0) had an average normalized UMI count of 5.86. Sub-cluster 1 (intermediate maturity) had an average of 3.36, and sub-cluster 2 (most immature) an average of 0.49. This increase in *CNTNAP2* expression with culture maturity was also observed in the *CNTNAP2* qPCR time-course described in **Chapter 4**.

The second highest expression of *CNTNAP2* is found in Cajal-Retzius neurons, which had a normalized UMI count of 3.59. The cortical hem had the third highest expression (2.73 normalized UMI), followed by the pre-thalamic eminence (2.30), the choroid plexus (2.04), and pre-thalamic progenitors (1.24). The remaining cell types - cortical progenitors and the pre-thalamus – both weakly expressed *CNTNAP2* with average expressions of 0.44 and 0.73 respectively. There was no significant difference in *CNTNAP2* expression between the sub-clusters of the pre-thalamic eminence or cortical progenitors.

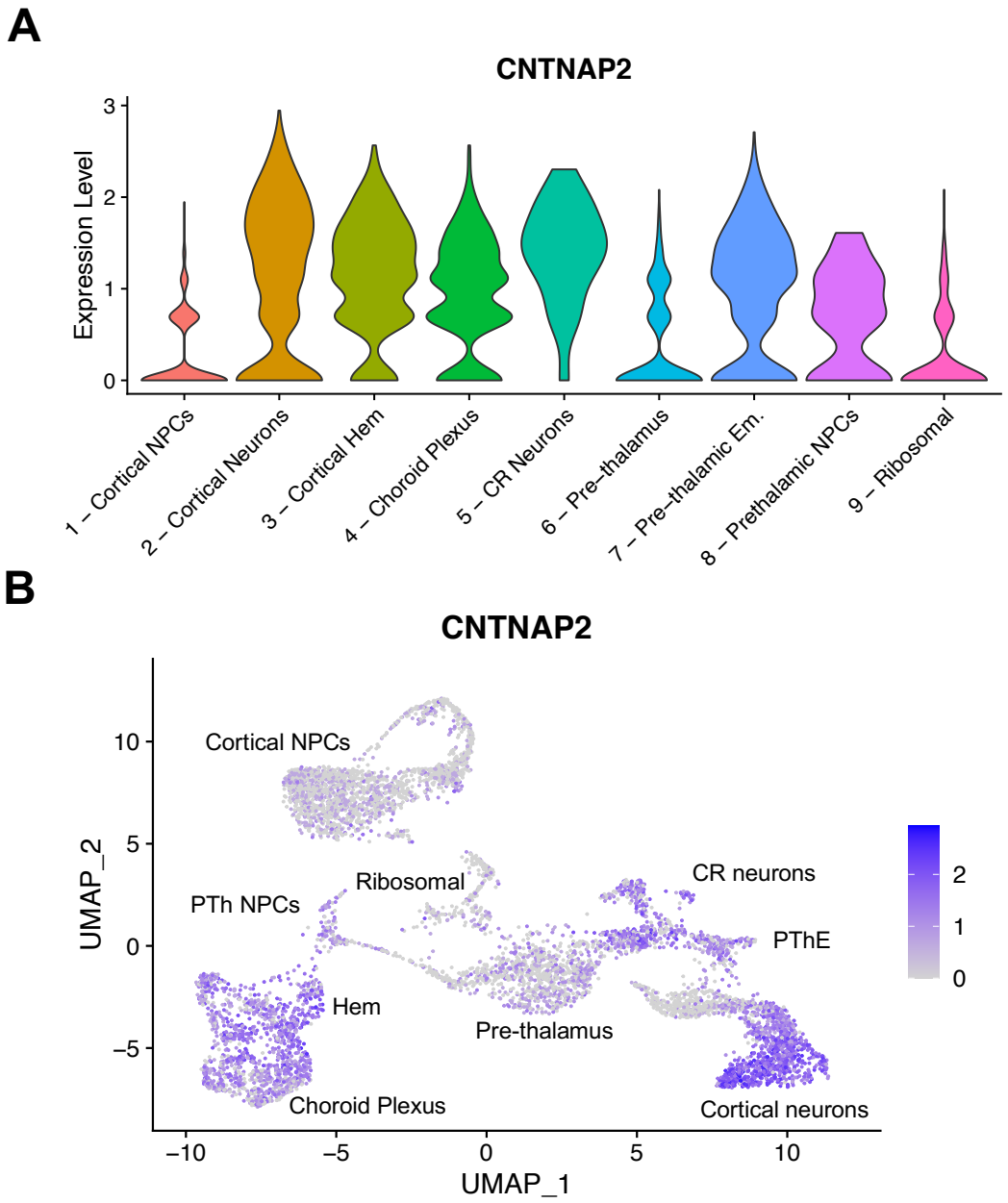


Figure 5.16. *CNTNAP2* expression across the nine cell clusters (WT cells only)

(A) Violin plot of *CNTNAP2* expression across the nine clusters (wild type cells only). Strongest expression is found in cortical neurons, and weakest in cortical progenitors. (B) UMAP plot of *CNTNAP2* expression in wild type cells. A clear gradient of expression can be observed in the cortical neuron cluster. This gradient correlates with maturity, with the most mature neurons (i.e. those with the strongest expression of neuronal markers like *MAPT*) showing the highest expression.

5.6.2 CNTNAP2 expression in WT versus KO cells

Having identified what cell types were present in the dataset, I first wanted to confirm that *CNTNAP2* expression was absent/reduced in the *CNTNAP2* KO cells of the scRNA-seq dataset. However, the average scaled expression of *CNTNAP2* between the two genotypes was not significantly different (Welch's *t*-test, $p = 0.96$) (**Figure 5.17**). Across all nine clusters, the expression of *CNTNAP2* in WT_1 was 1.7 average normalized UMIs, 1.7 UMIs in WT_2, 1.6 UMIs in KO_1, and 1.8 UMIs in KO_2. There are a number of explanations for why no apparent reduction was observed in the KO samples. Firstly, as outlined in **Chapter 3 – Methods**, the 10X scRNA-Seq platform sequences transcripts from their 3' end (i.e. they are captured by their poly-A tail). My CRISPR KO strategy targeted the 5' end of the gene, to increase the likelihood a truncated/non-functional protein would be made. This strategy targeted the canonical transcript, but was not able to affect the alternative transcripts that cluster at the 3' end of the gene (**Figure 2.02**, **Chapter 2**). It could be that there is still transcription of these 3' alternative transcripts, and this is what is being detected. Similarly, the reads assigned to *CNTNAP2* could actually belong to another gene that is transcribed in the opposite direction, and runs into the 3' end of *CNTNAP2*. For example, the Ensembl genome browser shows there are several non-coding RNA genes overlapping the *CNTNAP2* gene. It is possible that these are being detected and not *CNTNAP2*.

Another explanation is that a reduction in the *CNTNAP2* mRNA relies on the assumption of there being non-sense mediated decay. The presence of *CNTNAP2* expression in the KO samples could reflect that mRNA is still being produced in these cells (i.e. not decayed), but is not translated. As shown in **Figure 4.12** (**Chapter 4**), western blot confirms there is no *CNTNAP2* protein made – even at the C-terminus, where the antibody binds. Even if the loss of *CNTNAP2* isn't reflected in this scRNA-Seq dataset, I am still confident this is a full KO line given the Pac1, Sanger sequencing, MiSeq, and western blot assays all supporting it is (**Figure 4.11**). It is curious that the D50 qPCR I performed on neurons from the WT_1 and KO_1 inductions used in this scRNA-Seq experiment showed a ~75% reduction in RNA expression relative to WT.

However, all primers/reagents and PCR conditions were optimized, and the results replicated with multiple primers (targeting exons 3, 4, and 18).

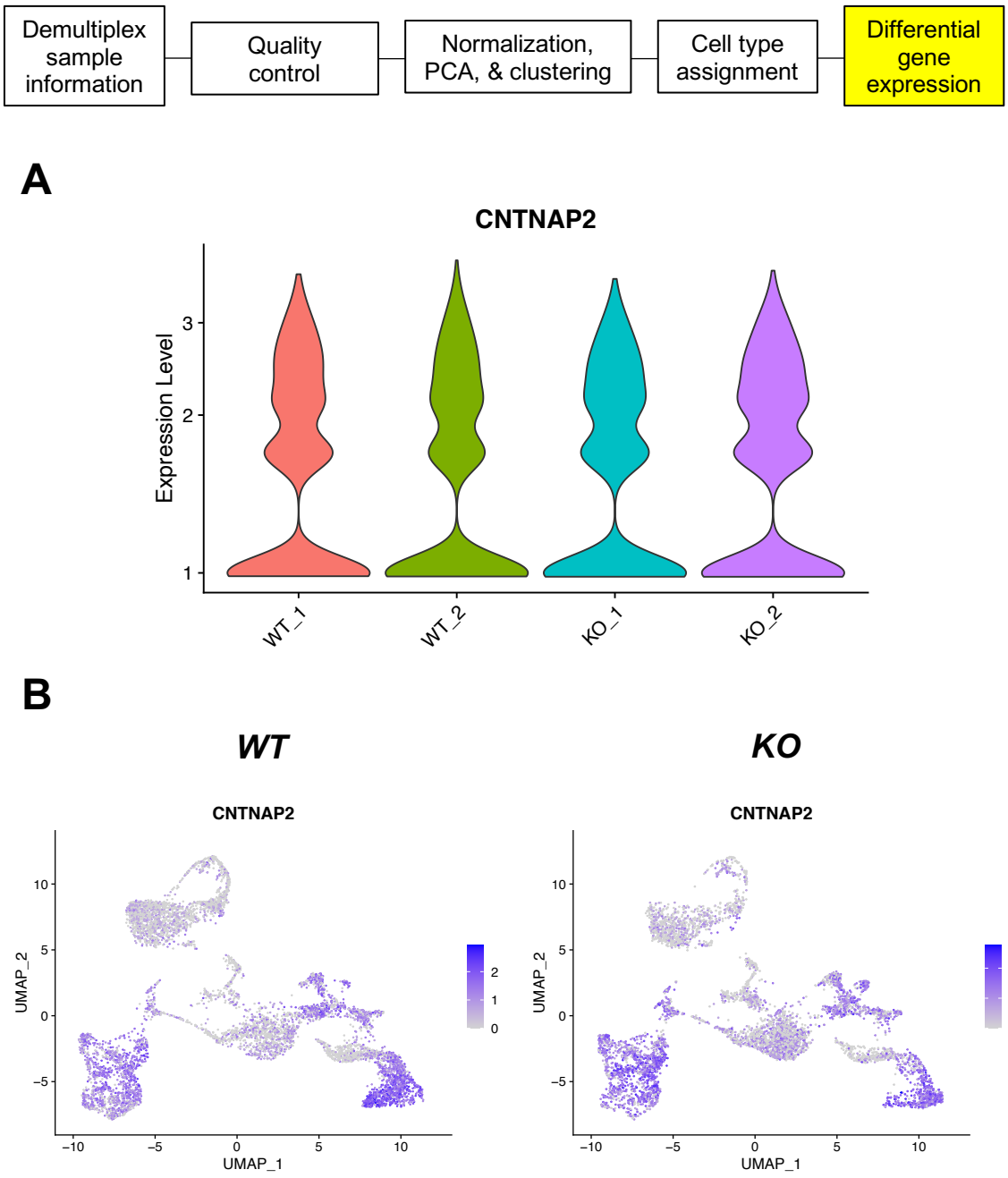


Figure 5.17. Reduction of *CNTNAP2* mRNA is not detected in KO cells

(A) Violin plot of *CNTNAP2* expression across the four samples (all cells plotted, irrespective of type). No significant difference in *CNTNAP2* expression is picked up between WT and KO samples. Potential reasons for this discrepancy are discussed in the chapter text. (B) UMAP plots of *CNTNAP2* expression in either wild type-only cells or knockout-only cells. These graphs reinforce that no difference in *CNTNAP2* was detected between the two genotypes, despite the unambiguous insertion of a stop codon and the absence of CASPR2 protein by western blot.

5.7 Overview of differential gene expression between WT and *CNTNAP2* KO forebrain cell types

To determine whether loss of *CNTNAP2* affects the expression of particular genes or pathways, I then used the *FindMarkers* function to identify genes that were differentially expressed (DE) between WT and KO cells in each of the clusters. Sub-clusters within the cortical neuron population (cluster 2) and pre-thalamic eminence population (cluster 7) were analyzed independently. Sub-clusters within the cortical progenitor population (cluster 1) were analyzed together given: 1) *CNTNAP2* is weakly expressed in cortical progenitors, and 2) *CNTNAP2* is not known to be involved in the cell cycle, which was the only meaningful distinction identified between the sub-clusters.

Genes with an average log-fold change (logFC) > 0.2 and a Bonferroni corrected p-value < 0.05 after are reported below. These thresholds were chosen to prioritize true changes that were strong enough to exert an effect. Fold-change was calculated as a WT/KO ratio, meaning negative logFC values are more highly expressed in KO and positive logFC values more highly expressed in WT. Across all cells in the dataset 89 DE events were detected, comprised of 68 individual genes (see the **Appendix** for a full list). 57 genes were upregulated in KO cells, and 32 downregulated. Mature cortical neurons had the largest number of DE genes amongst the 9 clusters (41 genes). However, most of the DE genes were only modestly different between WT and KO. The largest difference was observed in mature cortical neurons (i.e. cluster 2, sub-cluster 0), where the logFC of *CALB2* was -0.75 ($p = 1.12e\text{-}46$). This was followed closely by *DLK1* in choroid plexus progenitors, which was also the gene with the greatest down-regulation in KO (logFC $= 0.68$, $p = 1.03e\text{-}8$). Similarly, the most frequently dysregulated gene in KO cells was *CRABP1*, which was DE in four cell types: 1) cortical NPCs, 2) immature cortical neurons, 3) mature cortical neurons, and 4) pre-thalamic eminence. This was followed closely by *CALB2*, *DLK1*, *LHX2*, and *NRN1* – each of which was DE in three cell types (see **Figure 5.18**).

CNTNAP2 was detected as DE in only two cell types, mature cortical neurons and choroid plexus (i.e. two of the cell types that most strongly express *CNTNAP2*). In cortical neurons the average logFC was 0.27 ($p = 5.44e\text{-}12$), with expression in 99% of WT cells and 96% of KO cells. In the choroid plexus, the average logFC was -0.22 ($p = 0.005$, i.e. higher expression KO than WT). *CNTNAP2* was found in 80% of WT cells and 85% of KO cells.

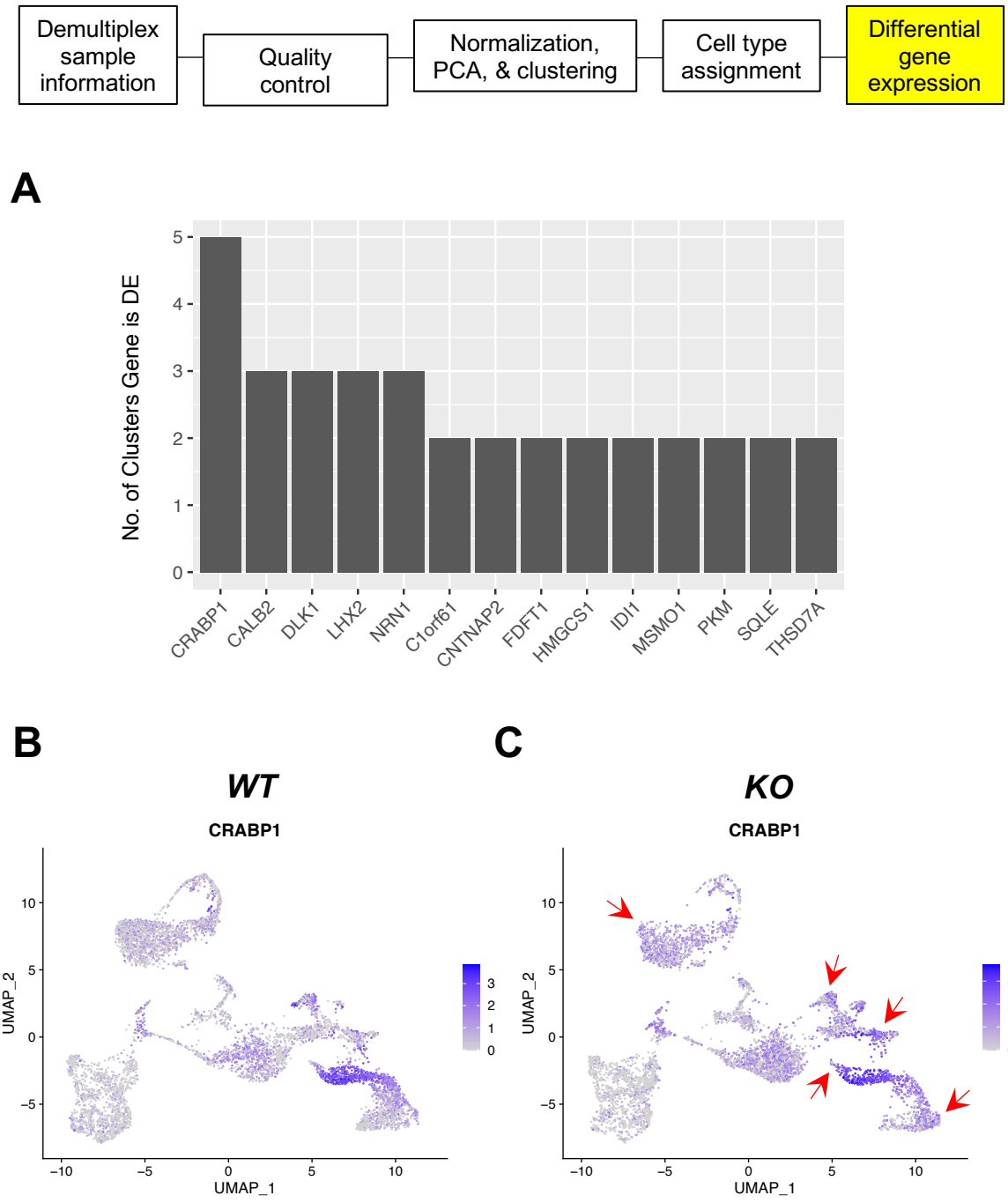


Figure 5.18. Repeatedly DE genes

(A) Plot of genes identified as differentially expressed in more than one cluster. Of the 68 individual DE genes, 14 were DE in 2 or more clusters. *CRABP1* was the most frequently DE gene, being significantly upregulated in KO cortical progenitors, KO immature cortical neurons, KO mature cortical neurons, and KO pre-thalamic eminence (both the *EMX1*$^{+}$ sub-cluster and *LHX5*$^{+}$ sub-cluster). (B) UMAP plots of *CRABP1* in only WT cells (left) and only KO cells (right). Red arrows point to clusters that *CRABP1* was upregulated in.

5.8 Analysis of DE genes between WT and *CNTNAP2* KO forebrain cultures

Having identified 68 DE genes in the dataset, I next sought to identify links between these genes and *CNTNAP2*. A literature search revealed 45 genes (or 66%) could be categorized into four main categories: **1) calcium signaling and/or neurotransmission genes** (seven), **2) neurite outgrowth genes** (five), **3) DNA regulation genes** (twenty-one), and **4) cholesterol metabolism genes** (eleven). My next step was to focus down this list by prioritizing genes involved in either **a) neurodevelopmental diseases** or **b) neurotransmission**. The first criterion would help highlight potential mechanism/pathways relevant for human disease and cognition. The second, could highlight pathways connected to the defects in synaptic activity previously noted in *Cntnap2* KO animal models. In total, 17 genes fit these criteria (25% of the total 68).

5.8.1 Calcium signaling and neurotransmission genes

Four calcium signaling genes were identified as DE, including: *CALB1*, *CALB2*, *TMEM97*, and *TRPM3*. A further three neurotransmission genes were also identified: *ATP1B1*, *NTS* , and *SYNPR*. Of these seven genes, five were DE in mature cortical neurons, including four up-regulated and one down-regulated [**up-regulated in KO:** *ATP1B1*, *CALB2*, *NTS*, and *SYNPR;* **down-regulated in KO:** *TMEM97*]. The remaining two genes, *CALB1* and *TRPM3*, were upregulated in KO pre-thalamus and downregulated in KO choroid plexus, respectively (see **Figure 5.19**, and **Table 5.01** for logFC, *p*-values, and percent cluster expression by each genotype).

These genes are of particular interest given the links between *CNTNAP2* and diseases of aberrant brain activity (e.g. epilepsy, schizophrenia). It is possible that the overexpression of neurotransmission genes and/or calcium signaling genes caused by *CNTNAP2* KO leads to neuronal hyperactivity and consequently neurological disease.

This hypothesis is strengthened by the fact that several of these DE genes have pre-established connections to such disorders.

For example, the *ATP1B1* gene encodes a subunit of the Na^+/K^+-ATPase that maintains the electrochemical gradients required for neuronal excitability (hence the gene's full name, Na^+/K^+-transporting ATPase subunit beta-1) (324). Defects in Na^+/K^+-ATPase activity have been implicated in epilepsy (325), bipolar disorder (326), depression (327), and Alzheimer's disease (328) – illnesses that *CNTNAP2* mutations have also been connected to. Similarly, Neurotensin (*NTS*) is a neuropeptide involved in neurotransmitter regulation (329). *NTS* has been repeatedly associated with disorders such as schizophrenia, autism, and addiction (330). This is also true of *TMEM97* (transmembrane protein 97), which encodes a sigma-2 orphan receptor (331). Moreover, neuroleptics (anti-psychotic medications) target both *TMEM97* and *NTS* proteins, further supporting that these genes are involved in cognitive processes. Finally, mutations in *TRPM3* (transient receptor potential cation channel, subfamily M, member 3) can cause intellectual disability and epilepsy (332).

The remaining three genes, *SYNPR, CALB1,* and *CALB2* don't have particularly strong associations with human disease, but were interesting candidates nonetheless. *SYNPR* (or synaptoporin) is a synaptic protein that is strongly expressed in the cortex (333). *CALB1* (calbindin) and *CALB2* (calretinin) are Ca^{+2}-binding proteins involved in calcium signaling and neurotransmission (334). Although *CALB2* is most commonly associated with interneurons, the gene is also expressed in glutamatergic neurons during development (335, 336). Crucially, there is some suggestion that *CALB2* plays a protective role against calcium overload/excitotoxicity (337). As such, the upregulation of *CALB2* could be connected to the frequent seizures in *CNTNAP2* patients.

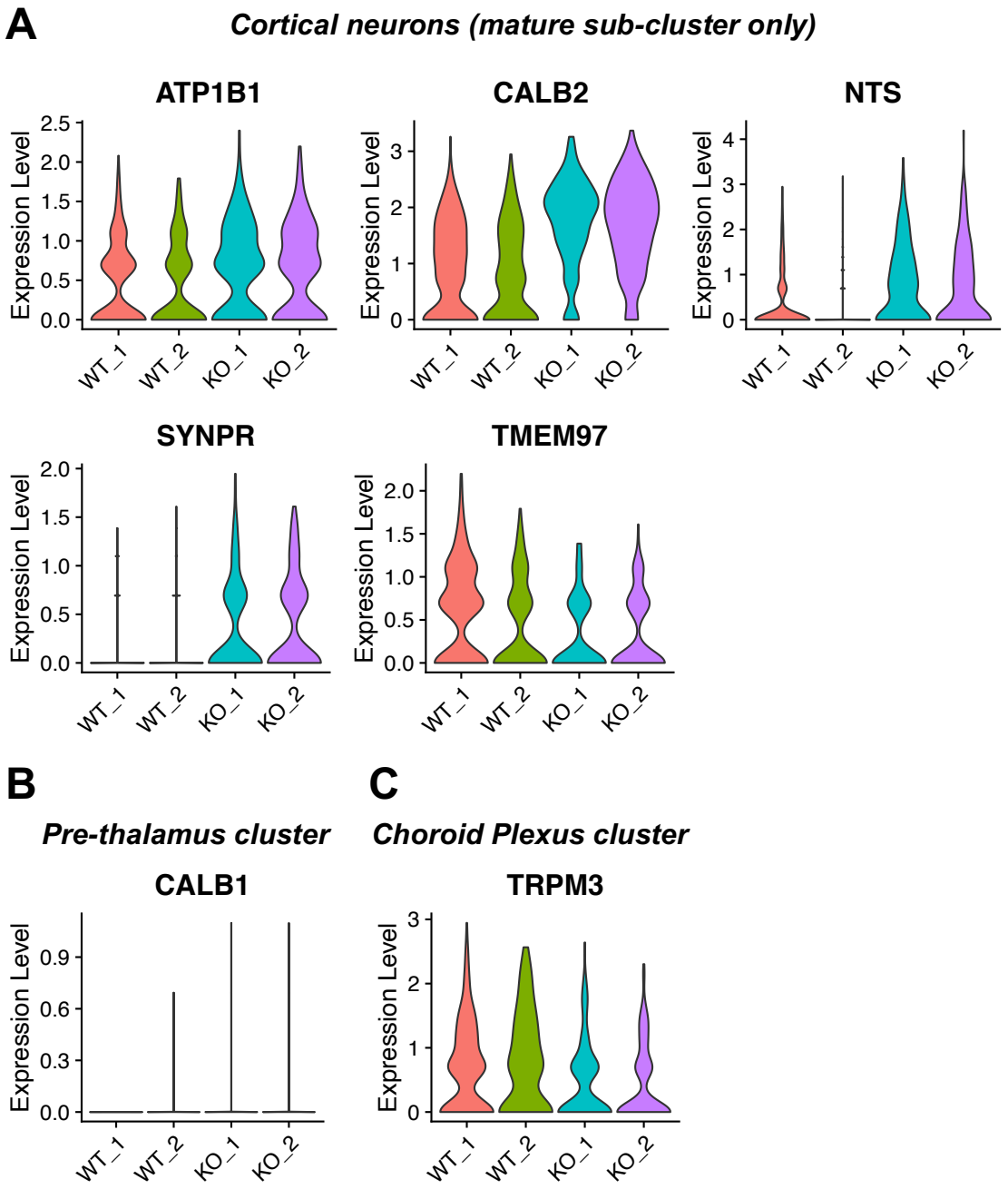


Figure 5.19. DE of calcium signaling and neurotransmission genes

Seven genes falling into this category were identified as differentially expressed (DE). (A) Five genes were DE in the mature cortical neurons sub-cluster, including four upregulated in KO neurons and one downregulated. Violin plots only include cells from this sub-cluster. (B) *CALB1* was identified as DE in the pre-thalamus cluster. However, the violin plot of pre-thalamic neurons shows expression is particularly weak in both WT and KO. (C) *TRPM3* was DE in the choroid plexus cluster – being upregulated in WT progenitors. Overall, five of the seven genes were upregulated in KO cells (and two downregulated). This finding potentially suggests an overactivity phenotype in the *CNTNAP2* KO cultures.

Gene	*Cluster DE*	*LogFC (WT/KO)*	*p-value*	*% WT cells*	*% KO cells*
ATP1B1	Mature cortical neurons	-0.211	0.0009	52	66
CALB1	Pre-thalamic NPCs	-0.61	0.005	6	28
CALB2	Mature cortical neurons	-0.75	1.12e-46	67	93
NTS	Mature cortical neurons	-0.59	1.32e-17	29	57
SYNPR	Mature cortical neurons	-0.28	7.88e-13	17	39
TMEM97	Mature cortical neurons	+0.28	9.86e-8	57	39
TRPM3	Choroid plexus	+0.35	2.47e-5	60	42

Table 5.01. DE of calcium signaling and neurotransmission genes

For each gene the cluster they were identified as differentially expressed (DE) in is reported, alongside the log fold change (logFC), corrected *p*-value, and the percentage of cells in the cluster expressing the gene (for both genotypes). LogFC is given for the WT/KO ratio, meaning positive fold changes represent a gene that is down-regulated in KO cells; negative fold changes, conversely, represent a gene that is up-regulated in KO cells.

5.8.2 Neurite outgrowth genes

Another interesting class to emerge from the DE analysis was neurite outgrowth genes. Five genes fell into this category: *NEFM*, *DOK5*, *TUBA1A*, *CXCR4*, and *NRN1*. *NEFM* was downregulated in KO pre-thalamus, and *DOK5* in KO immature cortical neurons. *TUBA1A*, *CXCR4*, and *NRN1* were all upregulated in KO choroid plexus, KO cortical progenitors, and KO cortical neurons respectively (see **Figure 5.20, Table 5.02**).

NRN1 (neuritin) was one of the most interesting in this class. *NRN1* encodes for an extracellular ligand important for neurite growth, synaptic maturation, and synaptic plasticity (338-341). *NRN1* expression increases in the brain with activity - and notably after seizures. Mutations in this gene have also been implicated in schizophrenia (342) and intellectual disability (343). *NRN1* was upregulated in all three sub-clusters of cortical neurons (i.e. immature, intermediate, and mature neurons). Once more, the higher levels of *NRN1* noted in the KO could be a sign of increased neuronal activity. It is unclear, however, whether this would be a cause or consequence of increased activity.

CXCR4 (C-X-C motif chemokine receptor 4) is another gene involved in neurite development that was DE in cortical lineage cells. In neural progenitors (NPCs), *CXCR4* accumulates in growth cones (protrusions that neurites grow from) to control neurite outgrowth and branching (344-347). *CXCR4* has been connected to multiple neurological disorders including epilepsy (348) and neurodegenerative diseases (349). Relatedly, mutations in *TUBA1A* (tubulin alpha 1a) have been identified in individuals with epilepsy, microcephaly, and lissencephaly (350, 351). This gene, which forms part of the neuron cytoskeleton, has also been shown to play a role in neurite outgrowth. In particular, iPSC-derived NPCs made from lissencephaly patients with *TUBA1A* mutations had reduced neurite extension (352). Finally, *DOK5* (docking protein 5) has been shown to promote axonal outgrowth in rodents (353).

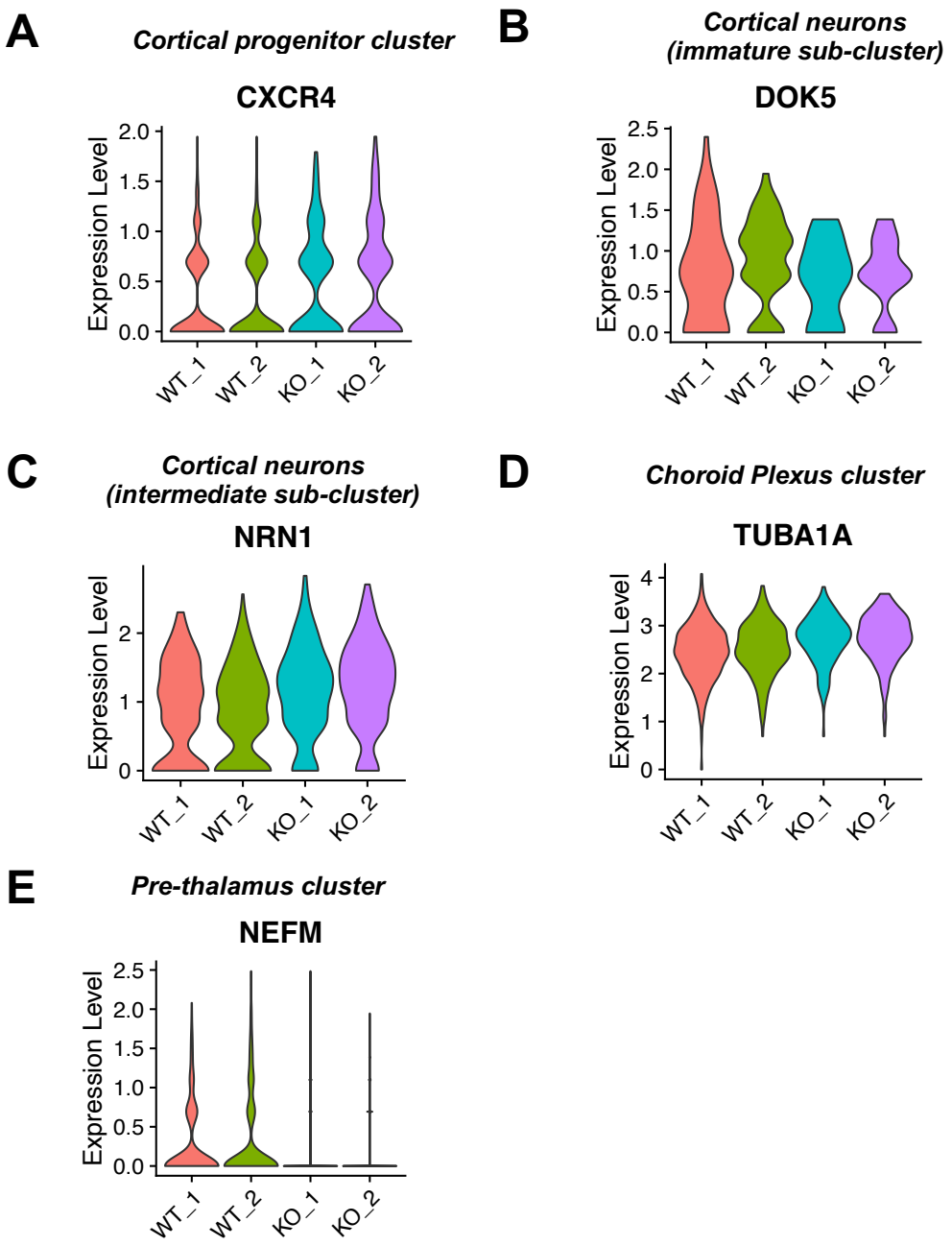


Figure 5.20. DE of neurite outgrowth genes

Five genes falling into this category were identified as differentially expressed (DE). Violin plots for each gene only include cells of the specified cluster/sub-cluster. (A) *CXCR4* was modestly upregulated in KO cortical progenitors. (B) Contrastingly, *DOK5* was downregulated in KO cortical neurons (immature sub-cluster). (C) *NRN1*, which had the greatest logFC of all five neurite outgrowth genes, was upregulated in cortical neurons (all three sub-clusters, with the greatest fold-change in intermediate neurons: logFC = -0.38). (D) *TUBA1A* was similarly upregulated in KO choroid plexus progenitors, while (E) *NEFM* was downregulated in KO pre-thalamic neurons. Overall, there is no trend in the direction of neurite outgrowth gene DE. As such, it is unclear which (if any) of these genes are truly dysregulated by loss of *CNTNAP2* contributing to KO phenotypes.

Gene	*Cluster DE*	*LogFC (WT/KO)*	*p-value*	*% WT cells*	*% KO cells*
CXCR4	Cortical progenitors	-0.22	2.96e-16	37	53
DOK5	Immature cortical neurons	+0.26	0.002	58	40
NEFM	Pre-thalamus	+0.23	3.61e-9	26	14
NRN1	Mature cortical neurons	-0.28	1.73e-6	73	88
TUBA1A	Choroid plexus	-0.22	3.84e-9	100	100

Table 5.02. DE of neurite outgrowth genes

For each gene the cluster they were identified as differentially expressed (DE) in is reported, alongside the log fold change (logFC), corrected *p*-value, and the percentage of cells in the cluster expressing the gene (for both genotypes). LogFC is given for the WT/KO ratio, meaning positive fold changes represent a gene that is down-regulated in KO cells; negative fold changes, conversely, represent a gene that is up-regulated in KO cells.

5.8.3 DNA regulation genes

NFIA and *NIFB* belong to the NF1 (nuclear factor 1) family of transcription factors. Deletions in both genes have been found in autistic individuals (215, 354, 355), and have also been implicated in bipolar disorder (356), intellectual disability (356, 357), and language impairment (357). Moreover, both genes interact with FOXP2 (which regulates *CNTNAP2* as previously discussed) (358). Repression of *NFIA* and *NFIB* in mice significantly reduces neurite outgrowth in cerebellar granule cells (359).

NOVA1 (neuro-oncological ventral antigen 1) is an RNA-binding protein involved in alternative splicing in the brain (360). Two studies have found mutations in *NOVA1* in children with neurodevelopmental disorders, including microcephaly, epilepsy, specific language impairment, and intellectual disability (361, 362). However, both studies involved large-scale deletions that affected a number of genes - whether *NOVA1* was specifically involved in generating these phenotypes remains unclear.

NEUROD2 (neuronal differentiation factor 2) is a member of the 'NEUROD' basic helix-loop-helix family of transcription factors that regulate neuronal differentiation (363). Mutations in *NEUROD2* have been identified in children with infantile epileptic encephalopathy (364). Moreover, the authors used CRISPR-Cas9 to reduce the expression of *NEUROD2* in tadpoles, which resulted in spontaneous seizures. In addition to epilepsy, *NEUROD2* mutations have also been associated with schizophrenia and autism (365).

Finally, *CELF4* (CUGBP elav-like family member 4) is another RNA-binding protein involved in the regulation of alternative splicing. In the SFARI Gene database, *CELF4* is listed as a high confidence autism candidate. Mutations in *CELF4* cause autism, seizures, and intellectual disability (354, 366-369).

NFIA, NIFIB, and *CELF4* were all up-regulated in KO mature cortical neurons, while *NEUROD2* was down-regulated. *NOVA1* on the other hand, was identified as up-regulated in KO pre-thalamic eminence. Please refer to **Figure 5.21** and **Table 5.03** for more detail.

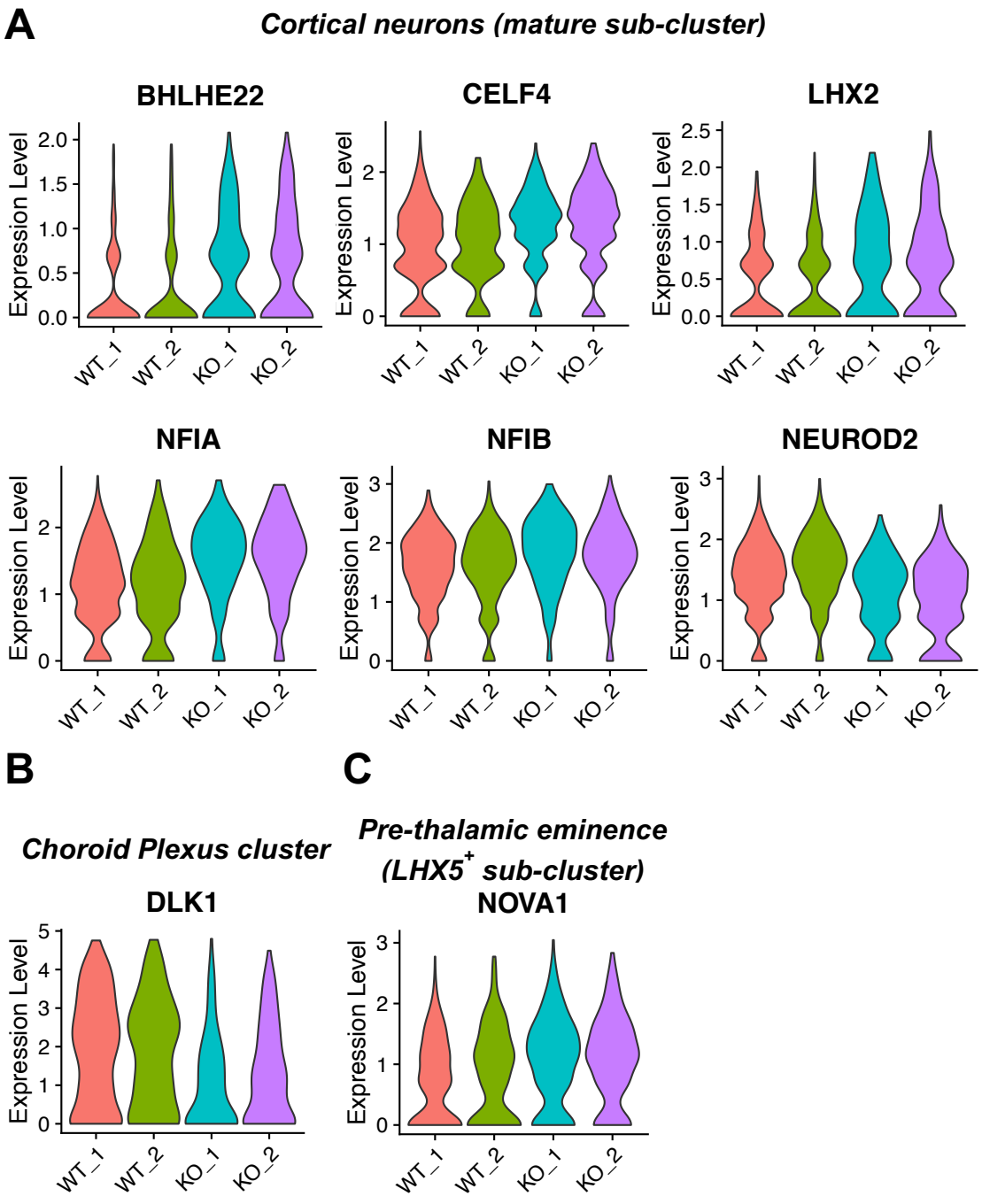


Figure 5.21. DE of DNA regulation genes

21 genes falling into this category were identified as differentially expressed (DE). Violin plots are shown for eight genes with the greatest connection to neurodevelopmental disorders (e.g. autism, schizophrenia). Plots for each gene only include cells of the specified cluster/sub-cluster. (A) Five example genes upregulated in KO mature cortical neurons, and one example gene downregulated. (B) *DLK1* was downregulated in choroid plexus progenitors, while (C) *NOVA1* was upregulated in the LHX5^{+} sub-cluster of pre-thalamic eminence.

Gene	*Cluster DE*	*LogFC (WT/KO)*	*p-value*	*% WT cells*	*% KO cells*
CELF4	Mature cortical neurons	-0.25	1.39e-10	84	93
LHX2	Mature cortical neurons	-0.28	0.001	55	65
NEUROD2	Mature cortical neurons	+0.37	3.95e-18	97	86
NFIA	Mature cortical neurons	-0.40	1.84e-24	87	94
NFIB	Mature cortical neurons	-0.23	2.91e-5	95	97
NOVA1	Pre-thalamic eminence (*LHX5*$^{+}$ sub-cluster)	-0.27	3.56e-5	53	78

Table 5.03. DE of DNA regulation genes (5 of 21 shown)

For each gene the cluster they were identified as differentially expressed (DE) in is reported, alongside the log fold change (logFC), corrected *p*-value, and the percentage of cells in the cluster expressing the gene (for both genotypes). LogFC is given for the WT/KO ratio, meaning positive fold changes represent a gene that is down-regulated in KO cells; negative fold changes, conversely, represent a gene that is up-regulated in KO cells.

5.8.4 Cholesterol metabolism genes

DHCR7 (7-dehydrocholesterol reductase) encodes for an enzyme that catalyzes the conversion of 7-dehydrocholesterol to cholesterol. It is ranked in the Simons Foundation Autism Research Initiative (SFARI) gene database as a high confidence autism candidate. Specifically, mutations in *DHCR7* cause a syndromic form of autism that is associated with Smith-Lemli-Opitz syndrome (SLOS) (370). SLOS is a rare autosomal recessive disorder that causes cognitive impairment, behavioral abnormalities and nervous system defects.

FDFT1 (farnesyl-diphosphate farnesyltransferase 1 – also known as squalene synthase) is a second gene involved in cholesterol synthesis. Individuals with mutations in *FDFT1* have developmental delay, intellectual disability, and dysmorphisms that resemble Smith-Lemli-Opitz syndrome (371). *FDFT1* was down-regulated in pre-thalamic eminence and mature cortical neurons, while *DHCR7* was only down-regulated in mature cortical neurons (**Figure 5.22, Table 5.04**).

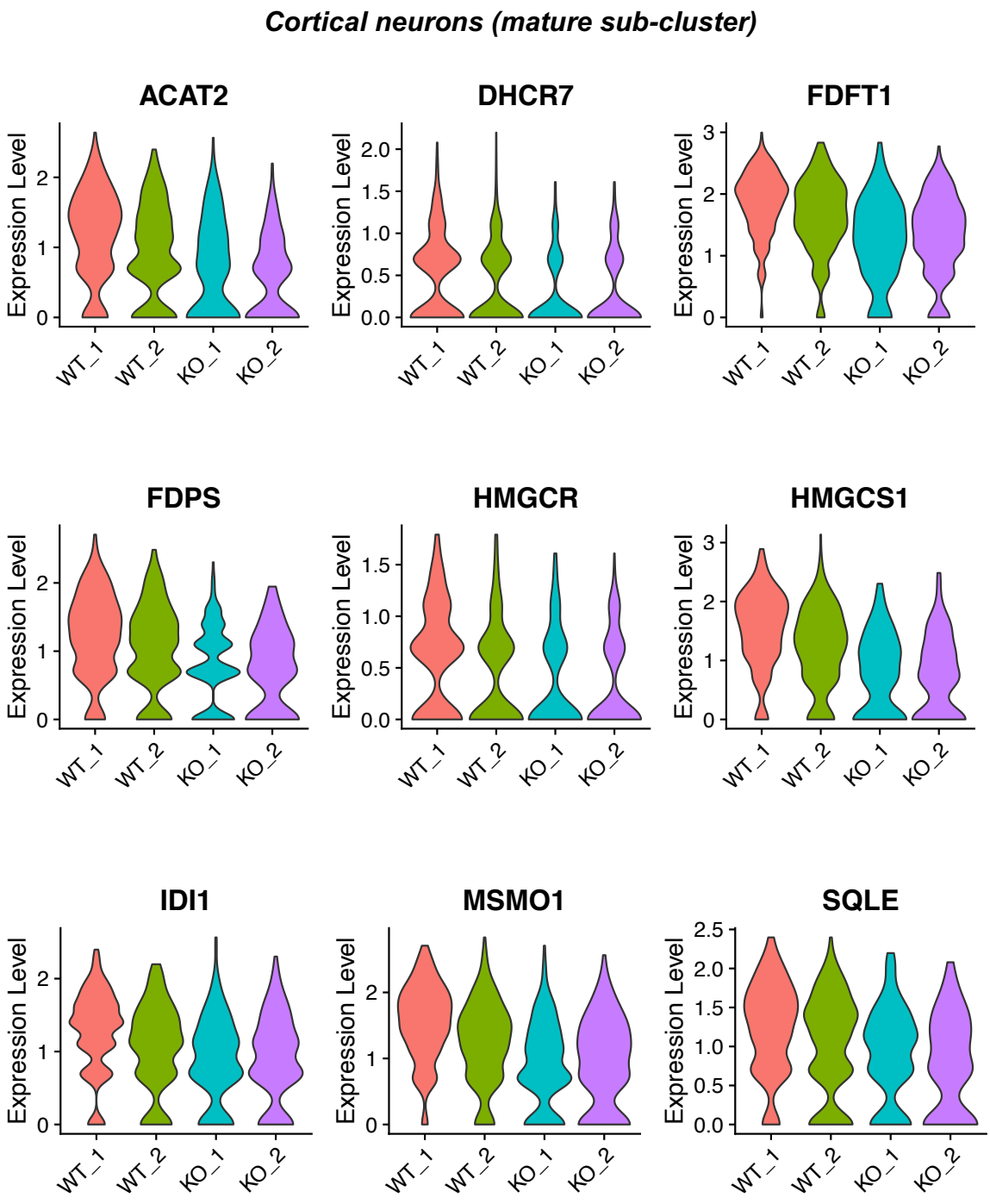


Figure 5.22. DE of cholesterol metabolism genes

11 genes falling into this category were identified as differentially expressed (DE). Violin plots for 9 of the 11 genes are shown. Plots only include cells from the mature cortical neuron sub-cluster. These genes are all members of the cholesterol synthesis pathway, and are all downregulated in KO neurons. While aberrant cholesterol metabolism has been connected to neurodevelopmental disorders (e.g. Smith-Lemli-Opitz syndrome), no connection between cholesterol synthesis and *CNTNAP2* or between these 11 genes and *CNTNAP2* is currently known.

Gene	*Cluster DE*	*LogFC (WT/KO)*	*p-value*	*% WT cells*	*% KO cells*
ACAT2	Mature cortical neurons	+0.46	5.39e-24	85	61
DHCR7	Mature cortical neurons	+0.23	1.18e-7	55	33
FDFT1	Mature cortical neurons	+0.43	8.43e-30	98	88
FDPS	Mature cortical neurons	+0.42	5.68e-19	87	72
HMGCR	Mature cortical neurons	+0.23	1.58e-7	60	39
HMGCS1	Mature cortical neurons	+0.63	4.68e-40	90	66
SQLE	Mature cortical neurons	+0.33	1.48e-13	88	72

Table 5.04. DE of cholesterol metabolism genes (7 of 11 shown)

For each gene the cluster they were identified as differentially expressed (DE) in is reported, alongside the log fold change (logFC), corrected *p*-value, and the percentage of cells in the cluster expressing the gene (for both genotypes). LogFC is given for the WT/KO ratio, meaning positive fold changes represent a gene that is down-regulated in KO cells; negative fold changes, conversely, represent a gene that is up-regulated in KO cells.

5.9 Discussion

In this chapter, scRNA-Seq was used to examine 1) WT *CNTNAP2* expression across a range of cortical cell types, 2) differences in culture composition between WT and *CNTNAP2* KO inductions, and 3) genes that are differentially expressed by *CNTNAP2* KO.

Firstly, with respect to WT *CNTNAP2* expression, I observed *CNTNAP2* was expressed in cortical neurons, cortical hem, choroid plexus, Cajal-Retzius neurons, and pre-thalamic eminence (**Figure 5.16**). It was also weakly expressed in cortical progenitors and pre-thalamus. As such, *CNTNAP2* is robustly expressed in post-mitotic neurons – but not limited to them. These findings agree with existing studies conducted on primary human brain tissue. For example, *in situ* hybridization (ISH) studies on human fetal brain, found *CNTNAP2* is expressed in the frontal and anterior temporal lobes, medial ganglionic eminence, striatum, and dorsal thalamus (**Figure 2.05, Chapter 2**) (200, 201). Similarly, a published scRNA-seq investigation of human fetal cortex showed *CNTNAP2* is expressed by NPCs, choroid plexus, excitatory neurons, and inhibitory neurons (post-conception weeks 6-37) (287). Although performing the experiment at D50 had many advantages (discussed in the introduction of this chapter), it also meant the variety of cell types present in the cultures was reduced. For instance, upper layer neurons and astrocytes only start forming around D80, meaning *CNTNAP2* expression by these cells could not be measured. Ideally, follow-up experiments should be conducted at later timepoints (e.g. D80) when more cell types are present.

The second set of results from this chapter pertained to the comparisons of WT and *CNTNAP2* KO cultures. Firstly, I identified that there was no significant loss of cell types between the two genotypes (**Figure 5.15**). As mentioned, there is considerable evidence in mice that *CNTNAP2* KO leads to a loss of cortical interneurons (245, 269, 271, 272). This has never been investigated in a human system, but could readily

explain phenotypes experienced by patients with *CNTNAP2* mutations (e.g. epilepsy, autism, schizophrenia, etc.) (43). Unfortunately, no cortical interneurons were identified in our scRNA-Seq dataset (in any of the inductions, for either genotype). It was therefore not possible to say whether there was a reduction or not. 'Ventralising' the cultures to produce cortical interneurons, for example by adding purmorphamine (a hedgehog signaling agonist), could help solve this issue in the future (173). A second reason to study interneurons pertains to the expression of *CNTNAP2* in the Allen Brain Atlas' single nucleus RNA-Seq (snRNA-Seq) dataset of adult middle temporal gyrus (MTG) (54). This dataset shows *CNTNAP2* is more strongly expressed in cortical interneurons than glutamatergic neurons. This finding was somewhat unexpected, given most studies have focused on *CNTNAP2* in excitatory neurons. It would therefore also be interesting to confirm whether *CNTNAP2* expression is higher in interneurons than excitatory neurons in the developing human cortex. From existing fetal scRNA-Seq data, it appears *CNTNAP2* expression is relatively equal between excitatory and inhibitory cortical neurons (287, 372). However, further studies are still needed.

The final results section from this chapter explored the DE genes between WT and *CNTNAP2* KO cultures. Four core categories of genes were detected as dysregulated: 1) calcium signaling and/or neurotransmission, 2) neurite outgrowth, 3) DNA regulation, and 4) cholesterol metabolism (**Figures 5.19-5.22**). Before analyzing these genes, it is important to first clarify a few points. None of the genes identified as DE are known to interact with *CNTNAP2* (according to the STRING protein interaction database (373)). Similarly, very few of the genes were predicted to interact with each other. The few interactions that were found, were mainly limited to the 11 genes of the cholesterol metabolism pathway. However, one could directly test these interactions with co-immunoprecipitation (co-IP) experiments. Co-IP experiments could also be extended to include *FOXP2, TCF4,* and *STOX1A* – the transcription factors that modulate *CNTNAP2* (as mentioned back in **Chapter 2**) (44). Repeating the experiment with another set of inductions will also help clarify noise from true effects. Additionally, most of the DE events were only modest in size. The

largest logFC was -0.75 ($p = 1.12e-46$, *CALB2* in mature cortical neurons). Again, repeating the experiment at a later timepoint, when the cultures are more functionally mature may lead to bigger expression differences.

That said, these results point to a few potential mechanisms of *CNTNAP2* function/dysfunction. Firstly, the finding that seven DE genes (~10% of the 68) were related to calcium signaling/neurotransmission was highly exciting. There is overwhelming evidence from rodent models that *CNTNAP2* is involved in these processes (34-36, 245, 261, 263). There are a few notable caveats to our results, however. Five of the seven were identified as increased in the KO cells, while two were downregulated [up-regulated in KO - **cortical neurons:** *ATP1B1*, *CALB2, NTS, SYNPR*; **pre-thalamus:** *CALB1;* down-regulated in KO – **cortical neurons:** *TMEM97*; **choroid plexus:** *TRPM3*]. It was curious that only seven specific genes were altered, and not others (particularly downstream binding partners). Once more, the only way to resolve their importance is to conduct additional experiments. Nevertheless, the finding that most of these genes were up-regulated in cortical neurons agrees with an excitation : inhibition imbalance. While it is not possible to definitively conclude whether these changes are a cause (or consequence) of altered activity, this question will be investigated in the next chapter.

The second category, neurite outgrowth genes, was also intriguing given the decreased neurite branching and/or decreased dendritic spine density in *CTNAP2* KO mice (**Figure 5.20**). As with the calcium signalling/neurotransmission genes, there was an inconsistency in the direction of change between WT and KO. This time, however, it was more severe, with about equal numbers up-regulated and down-regulated. For instance, *CXCR4* and *NRN1* were up-regulated in KO cells, however, *DOK5* was down-regulated. It is therefore not entirely clear, whether one can expect a reduction in neurite outgrowth, an increase, or no change at all. To address this, a complete neurite outgrowth assay will be presented in **Chapter 6**.

The final two categories, DNA regulation genes and cholesterol metabolism genes, were harder to interpret (**Figures 5.21-5.22**). As stated, several have been implicated in neurodevelopmental disorders, but none were known to directly interact with *CNTNAP2*. Similarly, it was unclear why eleven cholesterol metabolism genes were identified as DE (**Figure 5.22**). This result was unexpected given the absence of any published connections between *CNTNAP2* and cholesterol. These differences could be an artefact of the lines being in different stages of their metabolic cycles (i.e. unrelated to the loss of *CNTNAP2*), or could be a phenotype that is truly connected to the loss of the gene. Again, only further experimentation will be able to clarify this (e.g. additional scRNA-Seq studies, co-IPs, and so on).

Another approach to interpreting the DE genes is to compare them to other RNA-Seq datasets from *CNTNAP2* KO models. As a reminder, three datasets currently exist. The first, was conducted on cortical neurons derived from a schizophrenic patient with a heterozygous *CNTNAP2* mutation (239). The second, was conducted on forebrain organoids generated from three patients with a homozygous c.3709DelG mutation in *CNTNAP2* (295). Finally, the third, was on hippocampal tissue taken from *Cntnap2* KO mice (307). With respect to the first study (heterozygous mutation), only one gene, *C1ORF61*, was shared between the datasets. However, the direction of change was opposite to what was noted in our study (down-regulated in their patient cells versus up-regulated in our KO cortical hem and choroid plexus). A hypergeometric test to assess the degree of overlap between DE datasets was not significant ($p > 0.05$). The homozygous KO dataset, however, had six common DE genes with us. These included: *CXCR4, LHX2, LY6H, NFIA, NFIB*, and *NTS*. All of these shared the same direction apart from *NTS*, which was downregulated in their KO cultures, but up-regulated in our mature KO cortical neurons. This overlap with our scRNA-Seq dataset was found to be statistically significant ($p = 0.002$, hypergeometric test). With respect to the third dataset (KO mouse hippocampus), surprisingly not a single gene was shared with our list.

Overall, our results were most similar to the other human homozygous KO dataset. This makes logical sense and also suggests these shared DE genes are stronger candidates to consider. It is disappointing, however, that only six genes were shared between our experiments. That said, these studies are different in a number of key ways. These include (but are not limited to): species differences (for the mouse study), differences in forebrain differentiation methods, bulk versus single cell approaches, the precise mutation to the *CNTNAP2* locus, and/or the relative age of the cells examined. It is possible that these differences could explain the lack of overlap in our respective results.

Despite this lack of consistency, our current scRNA-Seq experiment suggests mechanisms for further investigation. Considering a number of the DE genes are implicated in neurodevelopmental disorders, they may converge with *CNTNAP2* at a common pathway. They may therefore point to clues on how to treat/prevent their associated diseases. These pathways could be further dissected by using bioinformatic tools like weighted gene co-expression network analysis (WGCNA), and other types of enrichment or pathway analyses. Of note, I performed preliminary gene ontology (GO) analyses with 'Enrichr' (374), but found no significant enrichment among differentially expressed genes after multiple hypothesis correction.

In view of the potential role for *CNTNAP2* in human evolution, I lastly wanted to compare the genes identified as DE in our KO cultures with genes DE between human and non-human primate cortex. To do so, I cross-referenced our list of genes with a dataset identifying DE genes between human and macaque primary fetal brain (post-conception weeks ~10-20), and human and chimpanzee PSC-derived forebrain organoids (3). Surprisingly, 26 of the 68 DE genes in our study were also on the list of DE genes between humans and non-human primates. This included several members of the cholesterol synthesis pathway *(FDFT1, ACAT2, FDPS, HMGCR, IDI1, INSIG1,* and *MSMO1)*, the DNA regulation cohort *(NFIA, NFIB, CELF4, LHX2, C1orf61, NELL2, SOX4, ID2, WLS,* and *ZIC2),* neurite outgrowth cohort *(NEFM* and *DOK5)*, and lastly, one member of the calcium signalling cohort *(TMEM97)*. It will be important to

confirm whether any of these genes interact with *CNTNAP2* (either directly or indirectly), which would suggest their relevant pathways could be involved in human cognitive function.

In the next chapter, I functionally test the WT and *CNTNAP2* KO forebrain cultures in a series of neurite outgrowth and culture activity assays. Overall, the results from this chapter suggest a potential phenotype of over-activity in the KO cells. However, this prediction needs to be validated with experimental data. There is also no clear picture on how neurite length/branching will be affected. The functional experiments of **Chapter 6** will aim to clarify these and other unknowns.

Chapter 6

Modelling *CNTNAP2* loss-of-function in human stem cell-derived forebrain systems

6.1 Introduction

Many studies have suggested *Cntnap2* plays a role in synapse formation and function in rodents (see **Chapter 2**). Cortical neurons from *Cntnap2* KO mice have reduced dendritic branching (35), neurite length (35, 202, 261), and dendritic spine density (34, 36, 263). In addition, loss of cortical interneurons (199, 269, 271, 272, 375), abnormalities in potassium channel clustering (193, 194), and aggregation of glutamate receptors (34, 203, 262) have also been described. However, it remains unclear whether the gene serves a similar function in humans. Given the association between *CNTNAP2* and disorders involving synaptic dysfunction – e.g. autism

spectrum disorder (ASD), intellectual disability (ID), and schizophrenia (discussed in **Chapter 1**) – there are initial grounds to expect it may.

By using the human *CNTNAP2* KO line reported in the previous chapter, these and other questions were investigated. This is crucial to not only understand the contribution of *CNTNAP2* to human-primate differences, but to also provide important clues to treat or prevent human disease due to loss of *CNTNAP2* function. Approximately 25 studies of *Cntnap2* KO or KD models are currently published (most of which are rodent or zebrafish (269)). Although most of these studies note some effect on synaptic function, many are conflicting in their accounts. For example, Anderson et al. (35) found that shRNA-mediated *Cntnap2* KD reduced neurite branching and length in mouse cortical cultures; this was later replicated by Canali et al. (261) in mouse KO neurons. In contrast, Varea et al. (34) and Lazaro et al. (263) both reported no difference in neurite length (from *in vitro* and *in vivo* mouse KO neurons, respectively). Additionally, several independent groups detected a significant loss of cortical interneurons in their KO models (199, 245, 269, 271, 272). Others, however, reported they found no effect of *Cntnap2* KO on interneuron abundance (194, 263, 376).

Although these papers vary in certain key ways - e.g. KO vs. KD, *in vitro* vs. *in vivo*, 2D-culture vs. organoid, and so on - these differences do not correlate with specific findings. In other words, there are clear discrepancies in proposed *Cntnap2* function, even when taking into account experimental conditions. Further work is therefore needed to clarify the exact function(s) of *CNTNAP2* in both humans and other species.

Aims of Chapter 6:

i) Investigate the effect(s) of *CNTNAP2* knockout in human cortical excitatory neurons on:
    a. Neurite length and branching

b. Dendritic spine density
c. Network-level electrical activity

ii) Connect findings to mechanisms for neurological disorders caused by *CNTNAP2* mutations

iii) Connect findings to human evolution and known intra-primate differences

6.2 Neurite length and branching analysis

6.2.1 Experimental design

In order to visualize individual neurons in a densely populated culture, I sparsely labeled cells with the **mNeonGreen** fluorescent protein. mNeonGreen fluoresces with an excitation/emission maxima at 506/517 nm; it is extremely bright, highly stable, and has low cytotoxicity (377). Using this protein enables fine neuronal structures like neurites and dendritic spines to be accurately quantified. mNeonGreen was driven by a neuron-specific promoter, **Synapsin-1**, allowing for further specificity in our labelling approach. Both WT and KO cells were labelled by nucleofection at D35, and plated at 150,000 cells/cm^2 into 5 wells of a 96-well imaging plate (full details reported in **Chapter 3**). Nucleofection was used to standardize the age of the neurons studied. Only post-mitotic neurons present at the time of nucleofection (D35) will contain the fluorescent protein. In progenitors, for example, the fluorescent signal will be diluted out with subsequent rounds of cell division.

Following a recovery period of one week, the cultures were imaged on an Opera Phenix confocal microscope with a 20X water-immersion objective (numerical aperture 1.0). 12 fields of view were acquired for each well, with 9 z-planes taken per image. Importantly, the Opera Phenix can be automated to scan a standard portion of every well, meaning no bias is introduced by the experimenter selecting neurons to image. As well, since mNeonGreen is extremely bright, no further amplification of its signal is needed by staining with an antibody. The same cultures can therefore be re-imaged live across several developmental timepoints.

For all live imaging sessions, the plate chamber of the Opera Phenix was set to 37°C / 5% CO_2 and allowed to fully acclimate. Imaging continued from D42 every 3-6 days for five timepoints until D61 (**Figure 6.01a** and **Figure 6.02**). Following image acquisition, neurite length and branching were measured with **NeuronStudio**, a semi-

automatic neuron reconstruction software (**Figure 6.01c**) (378). Three parameters were sought: 1) **average branch length** per neuron, 2) **average neuron length** (defined as the sum of all the branches), and 3) the **average number of branches** per neuron (**Figure 6.01b**). To reduce bias introduced in this stage, all neurons of suitable quality were analyzed (i.e. not a selected subset). Suitable quality was broadly defined as any neuron that was a) clear/unambiguous, b) isolated/not overlapping other cells, and c) fully within the imaging window. Any neurons that did not confidently fulfill these criteria were omitted.

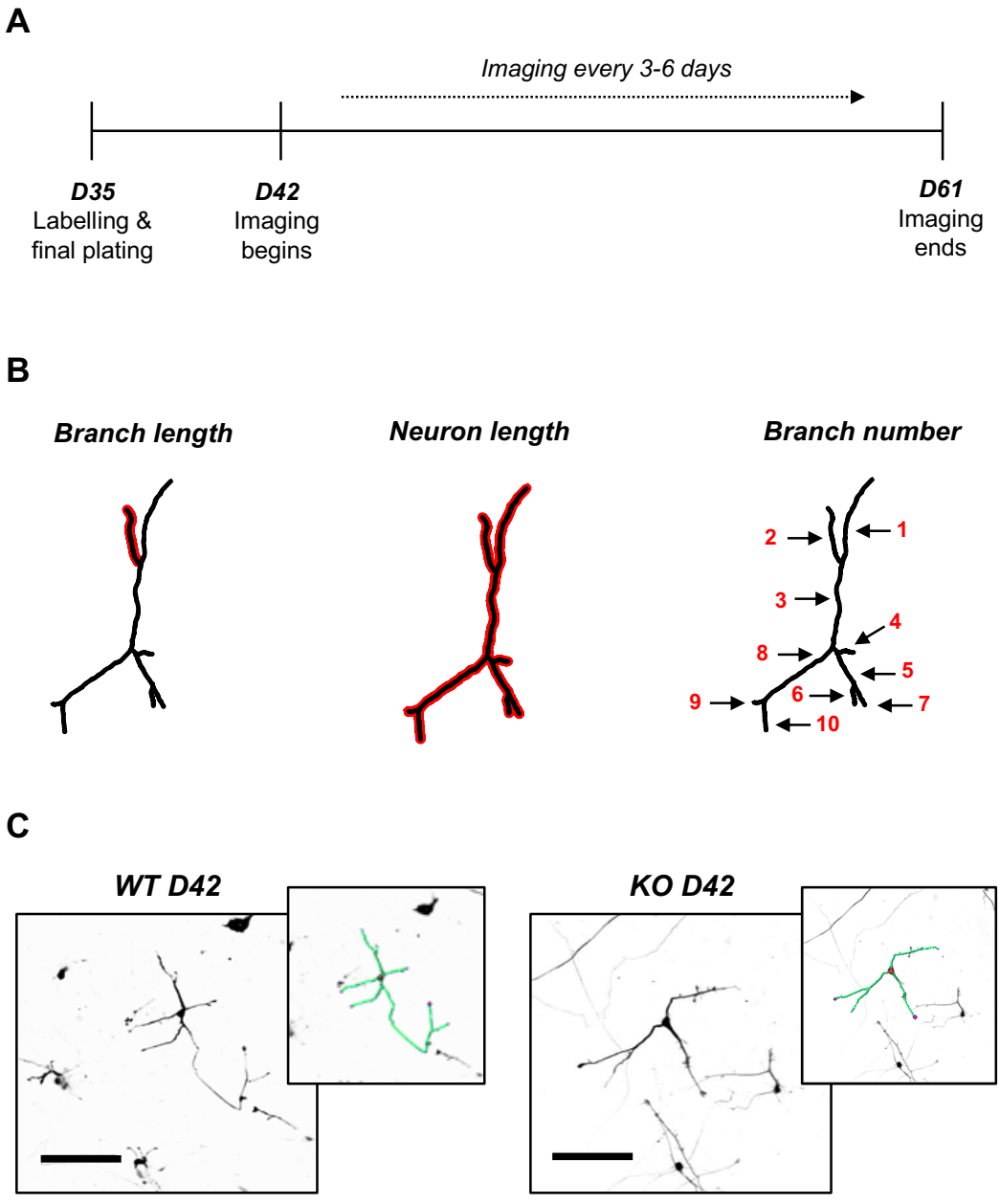


Figure 6.01. Neuron labelling strategy

(A) One induction each of WT and KO cells were nucleofected with mNeonGreen at D35. They were then immediately plated at 150,000 cells/cm^2. One week later, imaging commenced and continued every 3-6 days until D61. (B) Branch length is defined as the distance from a branch point to a branch tip, or to the next branch point. Total neuron length is the sum of all branches on a neuron, while branch number is simply the number of discrete segments composing a neuron. (C) Representative images of WT and KO D42 neurons taken on the Opera Phenix. Cell tracings from NeuronStudio accompany them on their right (red = soma, pink = branch tip). Scale bar measures 50 µm.

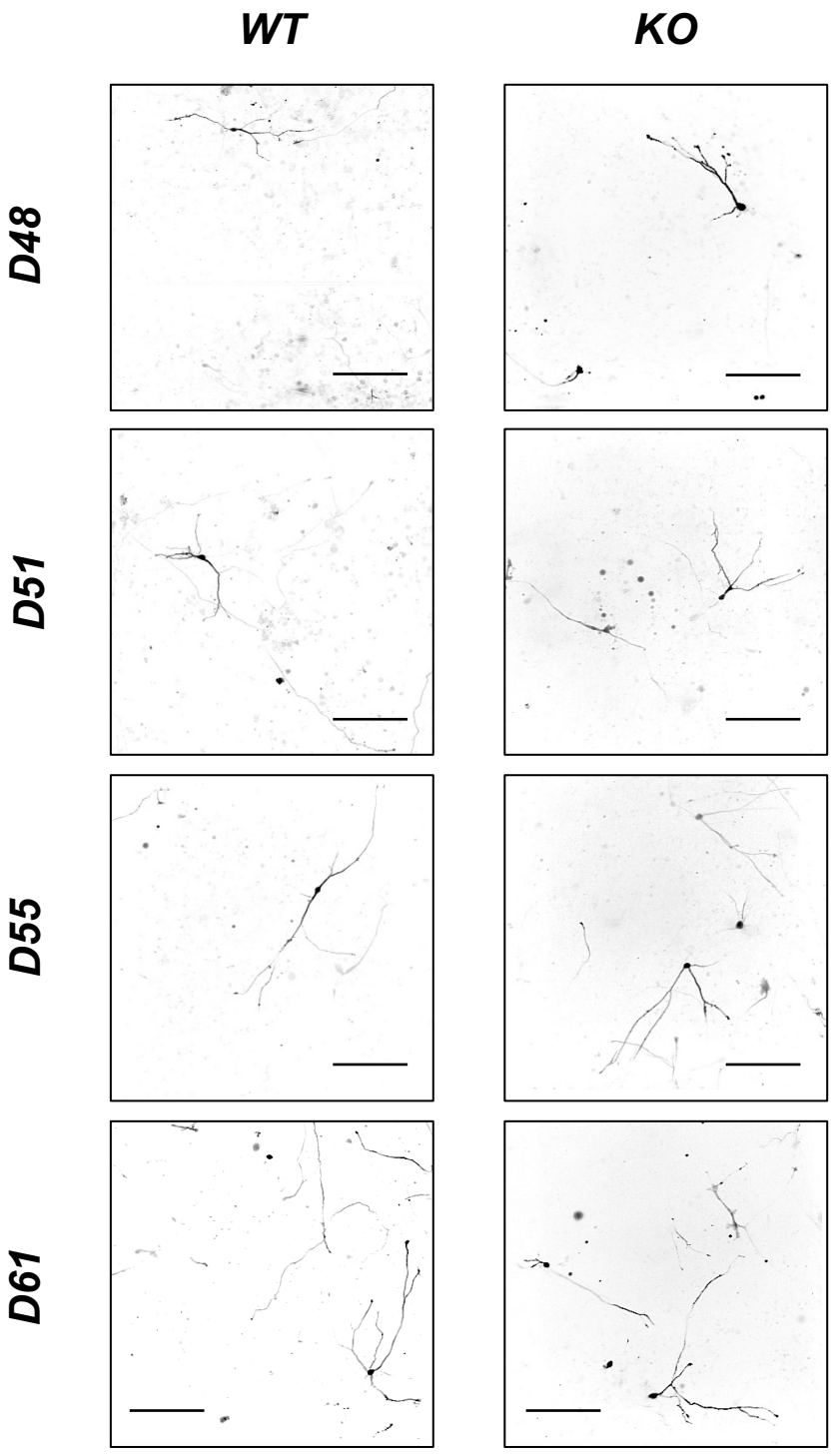


Figure 6.02. WT vs. KO sample live images

Live imaging time-course of WT and KO neurons. Four timepoints are shown, with each image representing a different field of view (i.e. a different neuron). Between 40 and 90 neurons were acquired for each timepoint, then semi-manually reconstructed using NeuronStudio software. Scale bar measures 100 μm.

6.2.2 Results of the neurite length and branching assay

A two-way repeated measures ANOVA was used to examine the effects of **1) *CNTNAP2* KO** and **2) culture age** on the three parameters discussed above. Firstly, for **average branch length** no significant difference was detected between WT and KO at any of the five timepoints [$F(1, 17) = 0.406$, $p = 0.533$] (see **Figure 6.03**). This was confirmed by post-hoc estimated marginal means comparisons, where $p > 0.05$ for all timepoints [**D42:** $p = 0.301$, **D48:** $p = 0.539$, **D51:** $p = 0.259$, **D55:** $p = 0.193$, and **D61:** $p = 0.129$]. Although ANOVA returned a significant signal for the effect of culture age on branch length [$F(4, 68) = 2.582$, $p = 0.045$], post-hoc analyses found no significant change between timepoints for WT, and only **D42 – D61** significant for KO ($p = 0.007$). Pearson's correlation for age and branch length was calculated as $r = 0.02$ for WT ($p = 0.81$) and $r = 0.289$ for KO ($p = 0.0003$).

When considering **total neuron length**, a significant effect was once more detected for culture age [$F(4, 68) = 7.91$, $p = 2.71\text{e-}5$], but not for *CNTNAP2* KO [$F(1, 17) = 0.079$, $p = 0.783$] (see **Figure 6.04**). Just as with branch length, no significant difference in total length was detected by pairwise comparison of WT and KO at each timepoint [**D42:** $p = 0.575$, **D48:** $p = 0.446$, **D51:** $p = 0.256$, **D55:** $p = 0.637$, and **D61:** $p = 0.483$]. Unlike for branch length, however, the relationship between total neuron length and culture age seemed stronger - Pearson's correlation was $r = 0.403$ for WT ($p = 1.31\text{e-}7$) and $r = 0.422$ for KO ($p = 2.61\text{e-}8$). Estimated marginal means detected several timepoints with different total lengths for both lines. Within WT, **D42 – D48:** $p = 0.0356$, **D42 – D51:** $p = 0.0311$, **D42 – D55:** $p = 8.65\text{e-}7$, and **D42 – D61:** $p = 2.83\text{e-}6$ were significantly changed. Within KO**, D42 – D51:** $p = 0.00239$, **D42 – D55:** $p = 0.0002$, and **D42 – D61:** $p = 0.0004$ were also calculated as significant.

Finally, for the **average number of branches per neuron**, no effect of *CNTNAP2* KO was detected [$F(1, 17) = 0.794$, $p = 0.385$] (see **Figure 6.05**). However, pairwise comparisons revealed KO cells had significantly reduced number of branches at **D55** ($p = 0.0254$). No difference in branching was detected between genotypes for the

remaining timepoints [**D42:** $p = 0.298$, **D48:** $p = 0.919$, **D51:** $p = 0.645$, and **D61:** $p = 0.0573$]. As with total neuron length (and unlike average branch length), culture age was found to impact branch number, with older cultures containing more branches per neuron [$F(4, 68) = 10.137$, $p = 1.70\text{e-}06$]. In WT cells branching increased significantly between **D42 – D51** ($p = 1.72\text{e-}3$), **D42 – D55** ($p = 4.97\text{-}13$), **D42 – D61** ($p = 2.70\text{e-}5$), **D48 – D55** ($p = 4.83\text{e-}4$), and **D51 – D55** ($p = 7.49\text{e-}3$). In KO cultures, significance was noted between **D42 – D51** ($p = 8.55\text{e-}3$) and **D42 – D55** ($p = 1.23\text{e-}4$). Similarly, Pearson's correlation between age and branching was $r = 0.466$ for WT ($p = 3.56\text{e-}10$) and $r = 0.229$ for KO ($p = 0.004$).

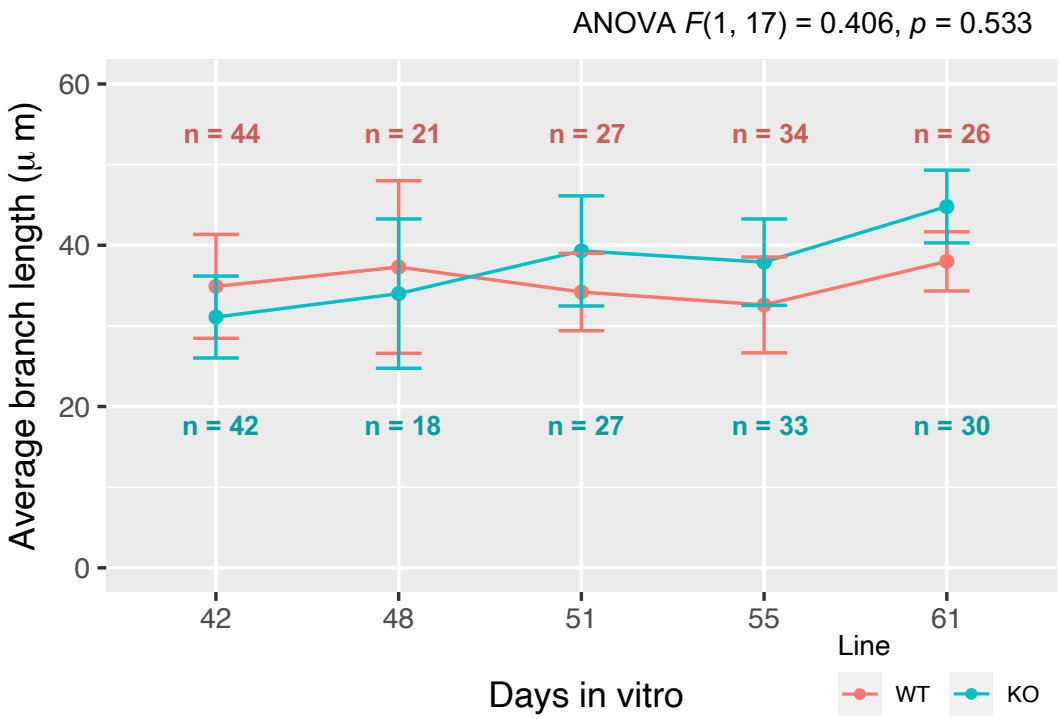


Figure 6.03. Average branch length per neuron

Two-way repeated measures ANOVA revealed no significant effect of *CNTNAP2* KO on average branch length per neuron (results printed above graph). Post-hoc pairwise comparisons subsequently confirmed this finding ($p > 0.05$ for estimated marginal means with Bonferroni correction). For the effect of culture age on branch length, ANOVA returned a significant signal [$F(4, 68) = 2.582$, $p = 0.045$]. However, no significant increase between timepoints was detected for WT, and only D42 – D61 was significant for KO ($p = 0.007$). Error bars represent 95% confidence intervals. Sample sizes (n) for each genotype/timepoint are shown in corresponding colours on the graph.

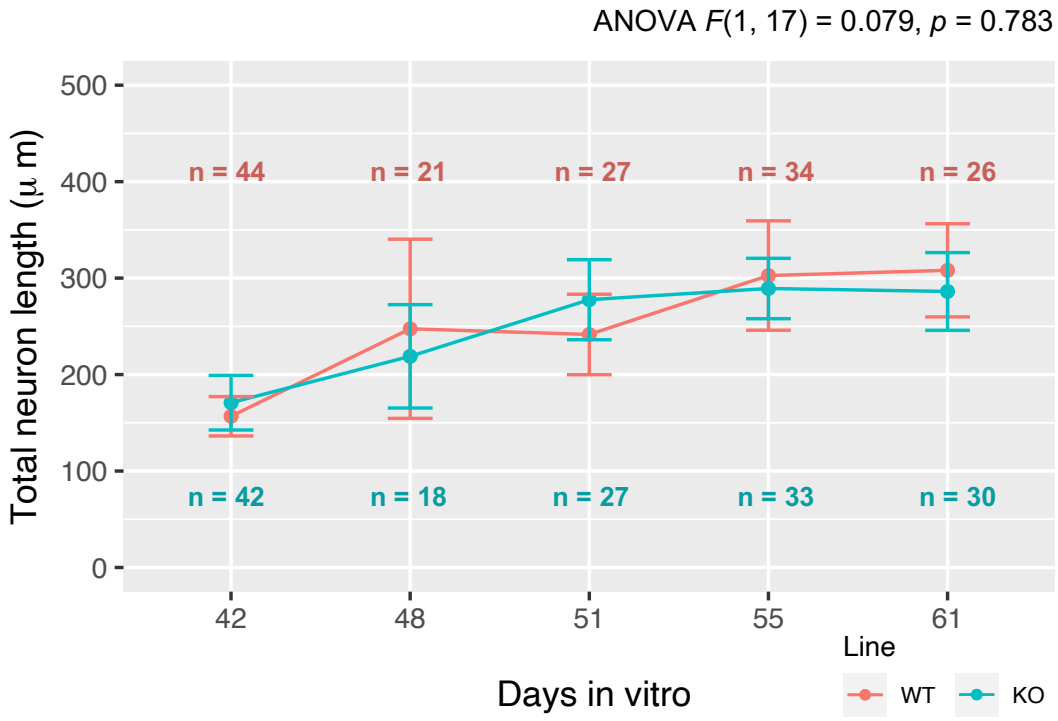


Figure 6.04. Average neuron length

Two-way repeated measures ANOVA revealed no significant effect of *CNTNAP2* KO on total neuron length (results printed above graph). Post-hoc pairwise comparisons subsequently confirmed this finding ($p > 0.05$ for estimated marginal means with Bonferroni correction). For the effect of culture age on neuron length, ANOVA returned a significant signal [$F(4, 68) = 7.91$, $p = 2.71\text{e-}5$]. Within WT, the following were significantly changed: D42 – D48 ($p = 0.04$), D42 – D51 ($p = 0.03$), D42 – D55 ($p = 8.7\text{e-}7$), and D42 – D61 ($p = 2.8\text{e-}6$). Within KO , the following were significant: D42 – D51 ($p = 0.002$), D42 – D55 ($p = 0.0002$), and D42 – D61 ($p = 0.0004$). Error bars represent 95% confidence intervals. Sample sizes (n) for each genotype/timepoint are shown in corresponding colours on the graph.

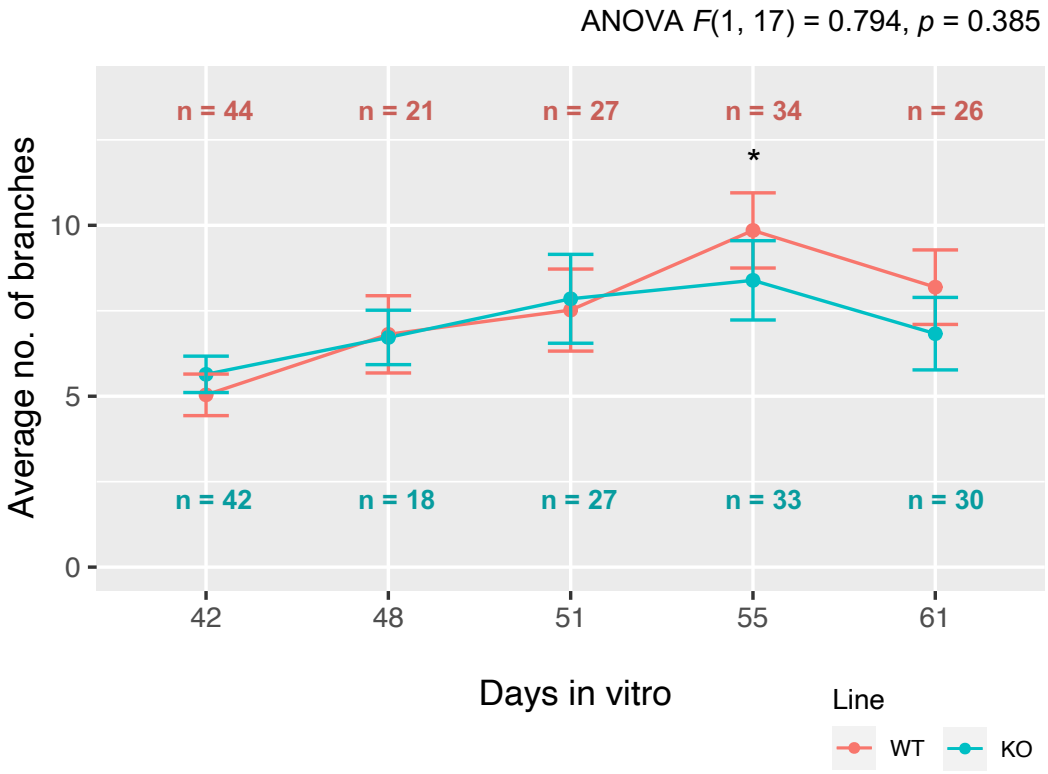


Figure 6.05. Average number of branches per neuron

Two-way repeated measures ANOVA revealed no significant effect of *CNTNAP2* KO on average number of branches per neuron. However, post-hoc pairwise comparisons revealed branching was significantly decreased in KO at D55 ($p = 0.03$, shown on graph) [estimated marginal means with Bonferroni correction]. For the effect of culture age on branch number, ANOVA returned a significant signal [$F(4, 68) = 10.14$, $p =$ 1.70e-06]. For WT, increases in branching were noted between D42 – D51 ($p = 1.7e-3$), D42 – D55 ($p = 4.97-13$), D42 – D61 ($p = 2.70e-5$), D48 – D55 ($p = 4.83e-4$) and D51 – D55 ($p = 7.49e-3$). In KO cultures, significance was noted only between D42 – D51 ($p =$ 8.55e-3) and D42 – D55 ($p = 1.23e-4$). Error bars represent 95% confidence intervals. Sample sizes (n) for each genotype/timepoint are shown in corresponding colours on the graph.

6.2.3 Neurite length and branching results in WT and KO co-cultures

Having observed that *CNTNAP2* KO may affect neurite branching, I repeated my initial experiment with certain key modifications. Since neurite and spine morphology are influenced by their environment (112, 379, 380), I wanted to better control for the effect of culture density/composition. Having each genotype plated separately made this impossible. Even if two cultures are final plated at the same cell density, differences in neural progenitor proliferation can make the final neuronal densities significantly unequal. To solve this, I adapted a protocol from Zaslavsky et al. (380) where WT and KO lines are differentially labelled and combined 1:1 into the same well. This design would also clarify whether differences between WT and KO neurons are **cell-autonomous** or **non-cell-autonomous** in nature.

The fluorescent protein **mKate2** was chosen for use alongside mNeonGreen, as it is also extremely bright and highly stable (fluorescing at 588/633 nm, far-red)(381). Different neuron-specific promoters were used for each protein. Synapsin-1 was kept in the mNeonGreen plasmid, while **CaMKIIa** was used for mKate2 (**Figure 6.06a**). Crucially, to avoid bias between the different proteins/promoters, two separate experiments were conducted: the first with WT cells labelled with mNeonGreen and KO cells with mKate2, and the second with labelling swapped between the genotypes.

As in the first set of experiments, cells were nucleofected with their respective fluorescent protein-expressing plasmids at D35, and plated at 150,000 cells/cm^2 (1:1 mix, 75,000 cells/cm^2 per genotype). Two inductions of WT and KO cultures were used (i.e. four inductions in total), with each induction filling 10 wells of a 96-well imaging plate. Following a recovery period of one week, the cultures were imaged on the Opera Phenix every 3-6 days from D42 to D66 (**Figure 6.07**). Empirically, the fluorescent signal weakened considerably by ~D60 - D65, which had also been noted in the previous experiment (albeit less strongly). This could explain the reduction in

branch number noted from D55 to D61 in experiment #1 (**Figure 6.05**). Because of this possible photobleaching, or difficulty in resolving neurites in dense cultures, all image acquisition was stopped at D66.

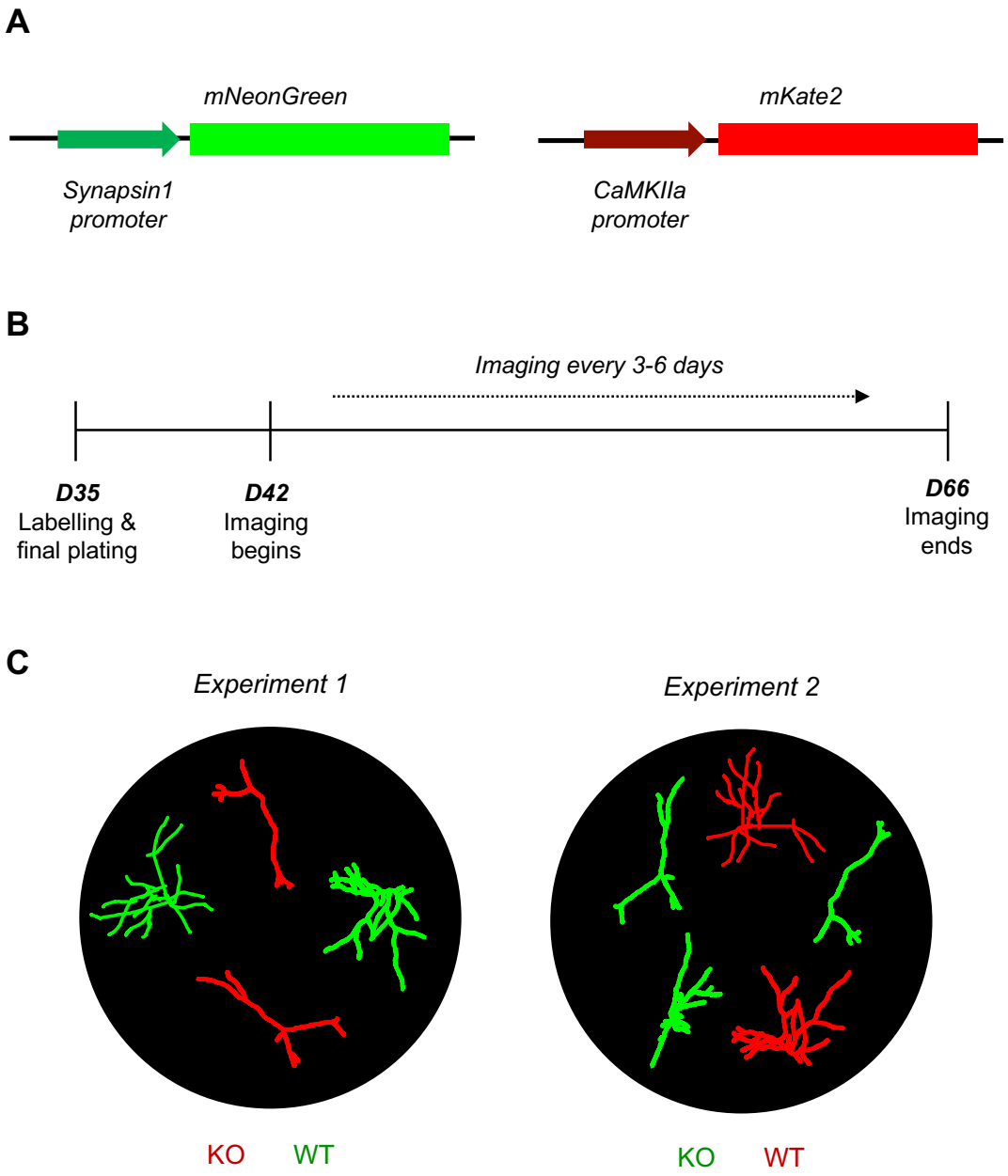


Figure 6.06. Neuron labelling strategy – mixed cultures

(A) WT and KO cells were differentially labelled then combined into shared wells at a density of 150,000 cells/cm^2. Synapsin-1::mNeonGreen and CaMKIIa::mKate2 (far-red) were used for their brightness, low cytotoxicity, and high stability. (B) Neurons were nucleofected with their assigned fluorescent protein at D35 before immediate plating. Following a one-week recovery period, cultures were live imaged every 3-6 days until D66. (C) To control for confounds caused by using different promoters/fluorescent proteins, two experiments were conducted with fluorescent proteins swapped between lines in the second iteration. For each experiment two inductions of WT and KO neurons were used.

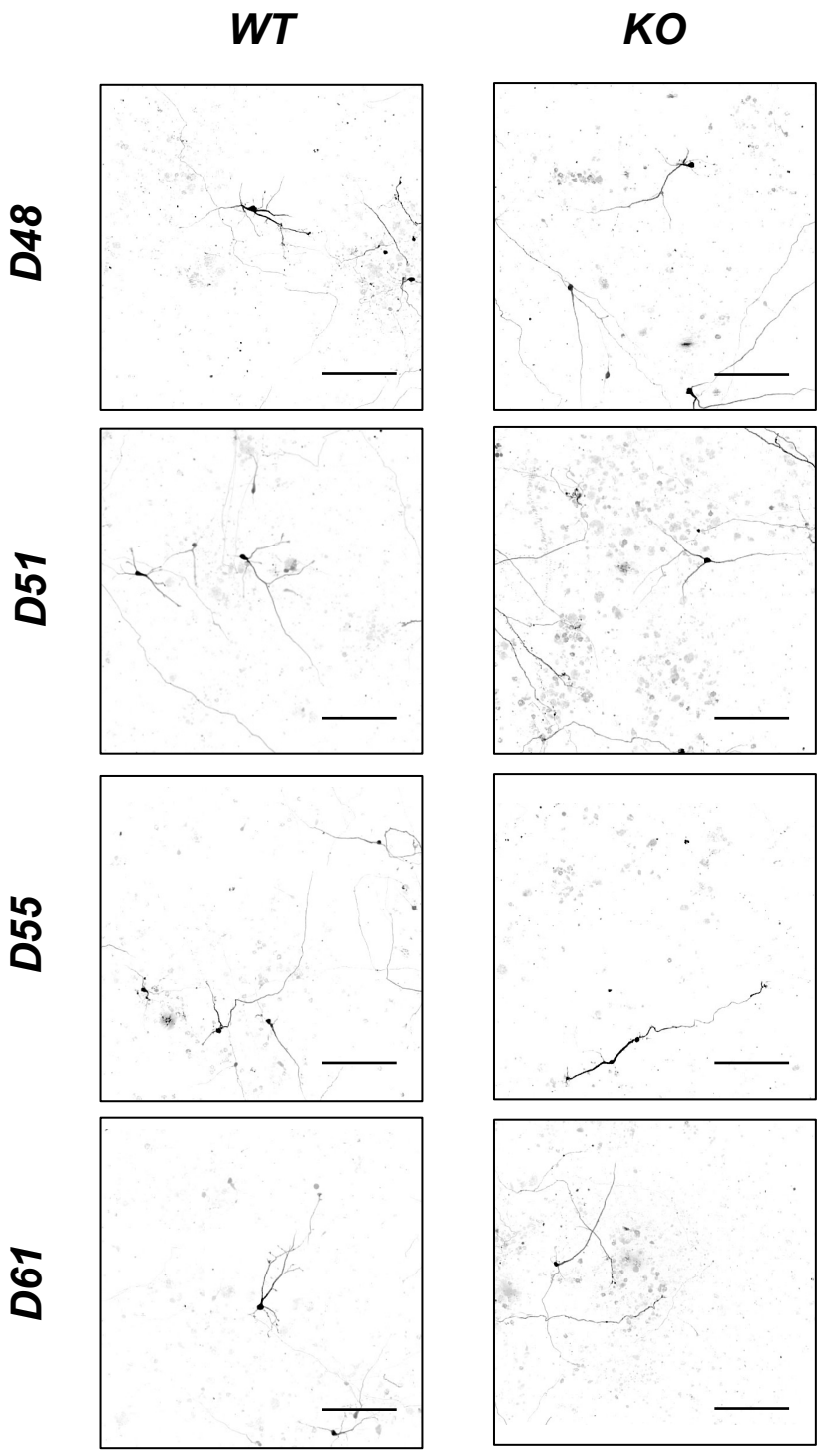


Figure 6.07. Mixed cultures - WT vs. KO sample live images

Live imaging time-course of WT and KO neurons (genotypes plated together in the same wells). Four timepoints are shown, with each image representing a different field of view (i.e. a different neuron). Between 70 – 250 neurons were acquired for each timepoint, then semi-manually reconstructed using NeuronStudio software. Scale bar measures 100 μm.

Similar to my first study, no significant difference was detected in **average branch length** between lines [$F(1, 35) = 1.139$, $p = 0.293$] (**Figure 6.08**). However, post-hoc pairwise comparisons found WT and KO did significantly differ at two timepoints after Bonferroni correction: **D48** ($p = 0.0171$) and **D60** ($p = 0.006$) [**D42:** $p = 0.999$, **D51:** $p = 0.345$, **D55:** $p = 0.557$, and **D66:** $p = 0.485$]. Moreover, branch length was detected as significantly changing with culture age [$F(3.67, 128.51) = 3.874$, $p = 0.002$]. Pairwise comparisons revealed significant increases between **D48 – D51** in WT neurons ($p = 0.015$) and between **D42 – D60** and **D55 – D60** in KO neurons ($p = 0.0367$ and $p = 5.28e\text{-}04$, respectively). Notably, a significant decrease in branch length was noted between **D60 – D66** in KO cells ($p = 0.019$). It is also important to note that changes to branch length – even when significant - were relatively modest and no real trend was observed over time. Pearson's correlation for branch length and age was calculated as $r = -0.01$ for WT ($p = 0.81$) and $r = 0.03$ for KO ($p = 0.48$). This implies branch length has not changed very much over the 24-day experiment, the same conclusion drawn from the first study.

Like with branch length, the ANOVA returned a significant effect on **total neuron length** over time [$F(3.63, 127) = 17.149$, $p = 1.33e\text{-}10$] (**Figure 6.09**). In WT cells, total length increased between **D42 – D51** ($p = 2.97e\text{-}09$), **D42 – D55** ($p = 5.18e\text{-}04$), **D48 – D51** ($p = 1.05e\text{-}08$), and **D48 – D55** ($p = 1.59e\text{-}03$). However, it was found to decrease between **D51 – D60** ($p = 2.86e\text{-}03$), **D51 – D66** ($p = 7.48e\text{-}06$), and **D55 – D66** ($p = 7.87e\text{-}03$). In the KO cells, total length increased between **D42 – D51** ($p = 8.01e\text{-}05$) and **D48 – D51** ($p = 2.59e\text{-}02$), and decreased between **D51 – D66** ($p = 1.53e\text{-}04$). Pearson's correlation was found to be $r = 0.07$ for WT ($p = 0.10$) and $r = -0.01$ for KO ($p = 0.78$). Unlike for branch length, however, *CNTNAP2* KO was also found to have a significant effect on total length [$F(1, 35) = 5.403$, $p = 2.60e\text{-}02$]. Pairwise comparisons detected KO cells had significantly reduced length versus WT at **D51** ($p = 0.0195$) and **D55** ($p = 0.00872$) [**D42:** $p = 0.518$, **D48:** $p = 0.381$, **D60:** $p = 0.167$, and **D66:** $p = 0.495$].

Regarding **average branch number**, both culture age [$F(1, 35) = 24.786$, $p =$ 1.71e-05] and *CNTNAP2* KO had a significant effect [$F(5, 175) = 14.541$, $p = 6.51e\text{-}12$] (**Figure 6.10**). Within WT neurons, estimated marginal means comparison found an increase in branch number between **D42 – D51** ($p =$ 5.32e-09), **D42 – D55** ($p =$ 1.90e-09), **D42 – D60** ($p = 0.0189$), **D48 – D51** ($p = 6.69e\text{-}04$), and **D48 – D55** ($p =$ 1.37e-04). Within KO, increases were found at **D42 – D51** ($p = 1.48e\text{-}07$), **D42 – D55** ($p = 1.31e\text{-}06$), **D48 – D51** ($p = 8.29e\text{-}03$), and **D48 – D55** ($p = 0.025$). However, significant decreases in branching were also detected in both lines: between **D55 – D60** in WT ($p = 1.688e\text{-}03$) and between **D51 – D60** ($p = 1.43e\text{-}07$), **D51 – D66** ($p =$ 3.64e-05), **D55 – D60** ($p = 1.55e\text{-}06$), and **D55 – D66** ($p = 1.17e\text{-}04$). Pearson's correlation was $r = 0.11$ for WT ($p = 0.01$) and $r =$ -0.04 for KO ($p = 0.37$). Comparing between lines revealed KO neurons had significantly fewer branches relative to WT at **D55** ($p = 0.0154$), **D60** ($p = 0.000005$), and **D66** ($p = 0.0414$) [**D42:** $p = 0.119$, **D48:** p $= 0.186$, and **D51:** $p = 0.0522$].

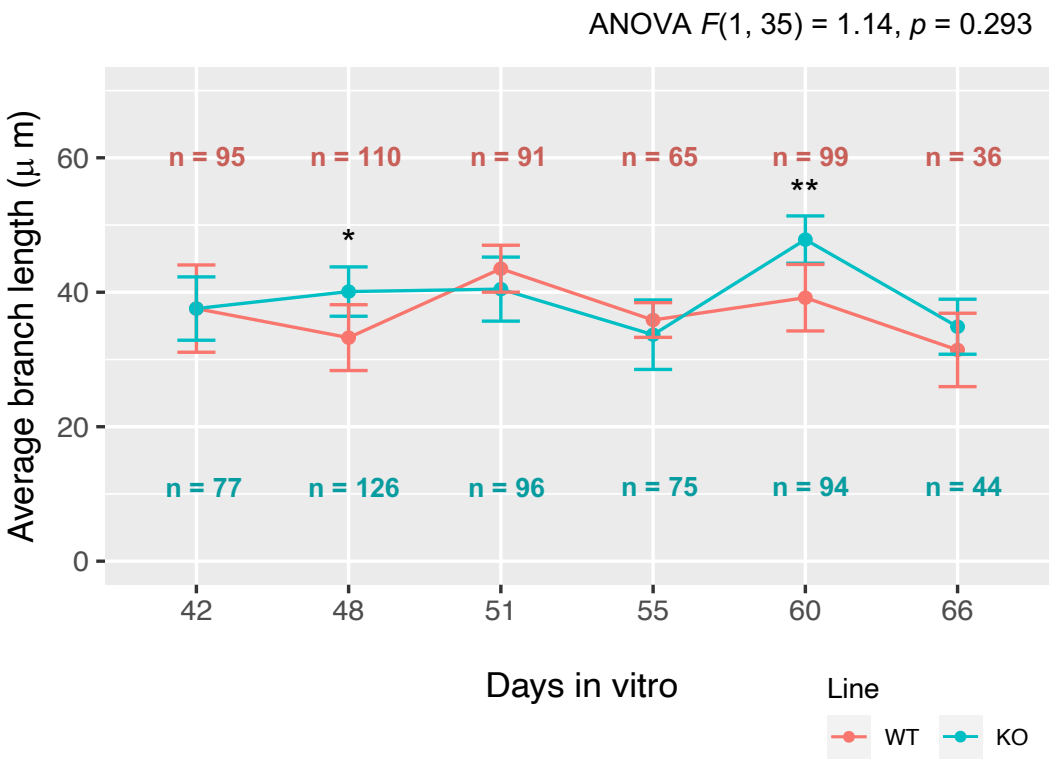


Figure 6.08. Average branch length per neuron (mixed cultures)

Two-way repeated measures ANOVA revealed no effect of *CNTNAP2* KO on average branch length per neuron. However, post-hoc analyses found length was significantly increased in KO relative to WT at D48 and D60 ($p = 0.0171$ and $p = 0.006$, respectively). For the effect of culture age on branch length, ANOVA returned a significant signal [$F(3.67, 128.51) = 3.874$, $p = 0.002$]. Pairwise comparisons detected the increase between D48 – D51 was significant in WT neurons ($p = 0.015$) and between D42 – D60 and D55 – D60 in KO neurons ($p = 0.0367$ and $p = 5.28e\text{-}04$). Notably, a significant decrease was found at D60 – D66 in KO cells ($p = 0.019$). Error bars represent 95% confidence intervals. Sample sizes (n) for each genotype/timepoint are shown in corresponding colours on the graph.

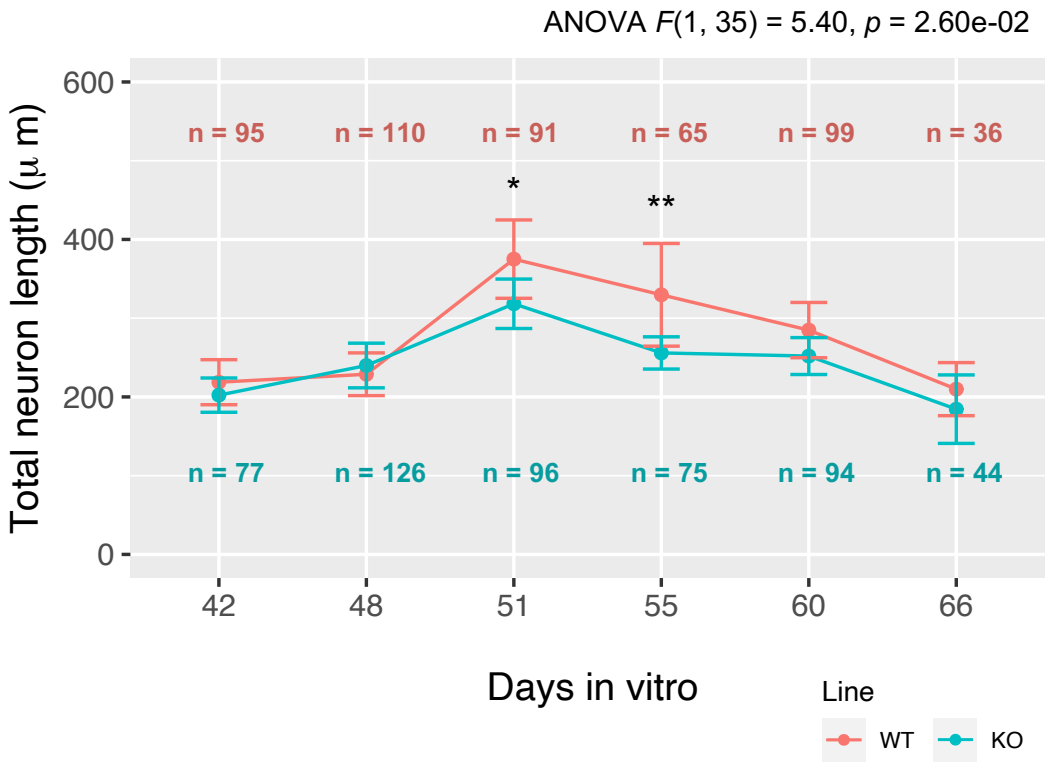


Figure 6.09. Average neuron length (mixed cultures)

Two-way repeated measures ANOVA revealed a significant effect of *CNTNAP2* KO on total neuron length. Post-hoc pairwise comparisons subsequently confirmed this finding, with KO cells showing reduced length at D51 ($p = 0.0195$) and D55 ($p = 0.00872$). For the effect of culture age on neuron length, ANOVA returned a significant signal [$F(3.63, 127) = 17.149$, $p = 1.33e\text{-}10$]. In WT cells, total length increased between D42 – D55 ($p = 5.18e\text{-}04$) and decreased between D51 – D66 ($p = 7.48e\text{-}06$). In KO cells, total length increased between D42 – D51 ($p = 8.01e\text{-}05$) and decreased between D51 – D66 ($p = 1.53e\text{-}04$). Error bars represent 95% confidence intervals. Sample sizes (n) for each genotype/timepoint are shown in corresponding colours on the graph.

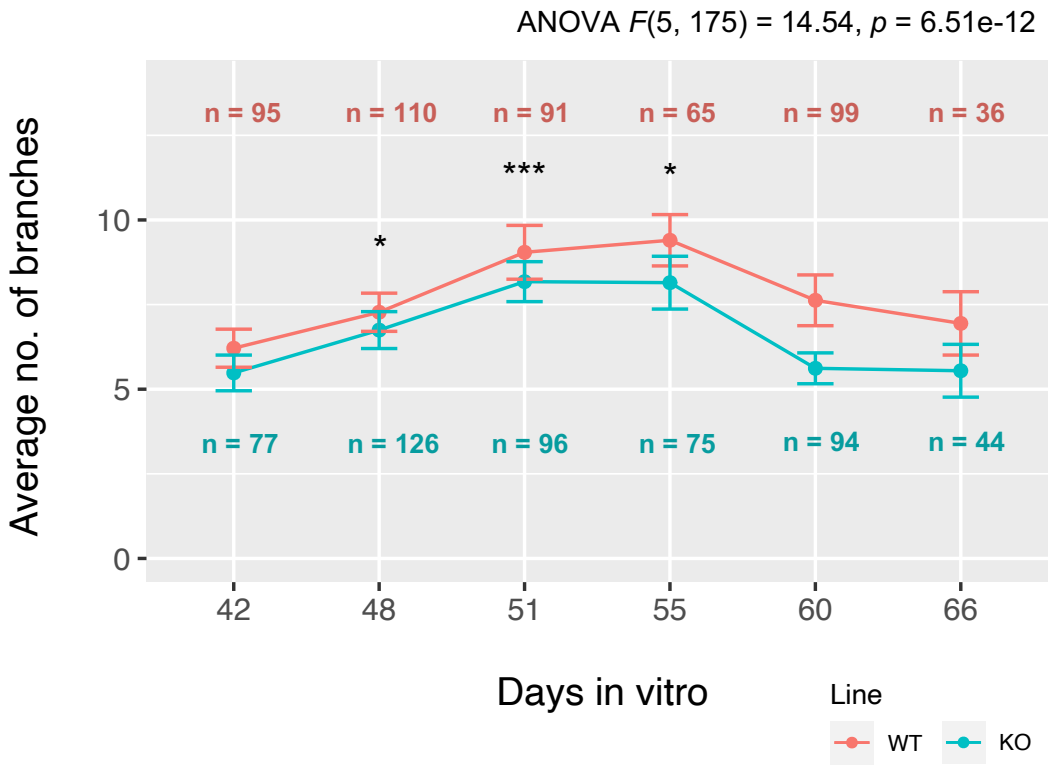


Figure 6.10. Average number of branches per neuron (mixed cultures)

Two-way repeated measures ANOVA revealed a significant effect of *CNTNAP2* KO on the average number of branches per neuron. Post-hoc pairwise comparisons revealed branching was significantly decreased in KO at D55 ($p = 0.015$), D60 ($p = 0.000005$), and D66 ($p = 0.0414$). For the effect of culture age on branch number, ANOVA also returned a significant signal [$F(4, 68) = 10.14$, $p = 1.70e\text{-}05$]. For WT, increases in branching occurred between D42 – D60 ($p = 0.019$), and decreases between D55 – D60 ($p = 1.69e\text{-}03$). In KO cultures, an increase was noted between D42 – D55 ($p = 1.31e\text{-}06$), and decrease between D51 – D66 ($p = 3.64e\text{-}05$). Error bars represent 95% confidence intervals. Sample sizes (n) for each genotype/timepoint are shown in corresponding colours on the graph.

6.2.4 Comparisons between experiments

Another key question to ask is whether the first experiment (separated genotypes) and the second experiment (mixed genotypes) are similar to each other. This will be important to understand whether the effects observed are cell autonomous or not. It will also inform on how consistent the measurements are between independent experiments.

A two-way ANOVA was performed to evaluate differences in **average branch length** between the two experiments (for each genotype). Since the first experiment ended at D61, only data from D42 – D60 were used from the second experiment (i.e. excluding D66). In the KO cultures the ANOVA revealed no significant difference in branch length between the two experiments [$F(1, 17) = 3.47$, $p = 0.08$]. The WT cultures, however, did return a marginally significant result [$F(1, 20) = 9.75$, $p = 0.04$]. Multiple pairwise comparisons (*t*-tests with Bonferroni correction) identified one timepoint as having significantly different branch lengths between WT cells in the unmixed and mixed cultures [**D55:** $p = 0.044$].

For **total neuron length,** the ANOVA observed a significant difference between experiments in WT cells [$F(1, 20) = 7.85$, $p = 0.01$]. No effect was detected between experiments for KO cells, however [$F(1, 17) = 0.43$, $p = 0.52$]. Multiple pair-wise comparisons identified WT cells had significant differences in total length at D42 ($p = 0.007$) and D51 ($p = 0.006$).

Finally, for **average number of branches** per neuron, ANOVA found no significant difference between experiments for WT [$F(1, 20) = 0.323$, $p = 0.58$]. However, it did return a significant result for KO cells [$F(1, 17) = 5.79$, $p = 0.028$]. Post-hoc analyses revealed branch number was significantly different across experiments in KO D60/61 cells ($p = 0.016$). Possible explanations for these differences in measurements between the experiments will be addressed in **Section 6.5 (Discussion)**.

6.3 Dendritic spine density

Accurately measuring dendritic spines required a few key modifications to my existing workflow. Since higher quality images would be needed (as small structures such as spines are not well resolved on the high-content Opera system), I switched from the Opera Phenix to a standard (i.e. not high-throughput) inverted confocal microscope. I also changed from a 20X/1.0 NA objective to a 60X/1.35 NA objective to further increase magnification/resolution. This change came at two notable costs: 1) imaging speed would be significantly decreased, meaning fewer neurons could be captured at a single time; and 2) I would now be manually selecting neurons to acquire, rather than automating the process (as done with the Opera Phenix). To reduce the possibility of bias, all imaging - and subsequent analysis - were conducted blind. As well, due to the increased imaging time, a decision was made to fix the cultures, rather than use live-imaging . By doing so, image acquisition could be extended over several days without the threat of spine dynamics significantly changing.

In order to capture a range of stages of neuronal maturity, cells were fixed at two timepoints: **1) D50**, around when spines first appear in the Shi et al. cultures (186), and **2) D75**, when spines would be more developed. Unfortunately, due to the COVID-19 pandemic, this experiment was not fully completed by the time of the lab's shutdown. One induction of WT and KO neurons was imaged, consisting of ~40 cells per timepoint. Both lines were nucleofected with mNeonGreen at D35 and plated separately into a 96-well imaging plate at 150,000 cells/cm^2. Ideally, this experiment would be repeated again with a greater sample size, mixing of WT and KO lines, and multiple independent inductions. Following image acquisition, NeuronStudio was used to count dendritic spines and measure branch length. **Spine density per neuron** was calculated as the average of the spine densities of each branch of the neuron. Examples of raw images and their reconstructions are shown in **Figure 6.11**.

A two-way repeated measures ANOVA was used to compare average spine density per neuron in WT and KO lines at each timepoint. Both *CNTNAP2* KO and culture age were found to have no significant effect on dendritic spine density [**KO:** $F(1, 16) = 1.31e\text{-}05$, $p = 0.997$; **culture age:** $F(1, 16) = 3.00e\text{-}03$, $p = 0.958$] (see **Figure 6.12**). Post-hoc pairwise comparisons showed similar results: at both **D50** and **D75** there was no difference in WT versus KO spine density ($p = 0.834$ and $p = 0.914$, respectively), and within each line, culture age also did not have a significant effect (WT: $p = 0.830$ and KO: $p = 0.519$).

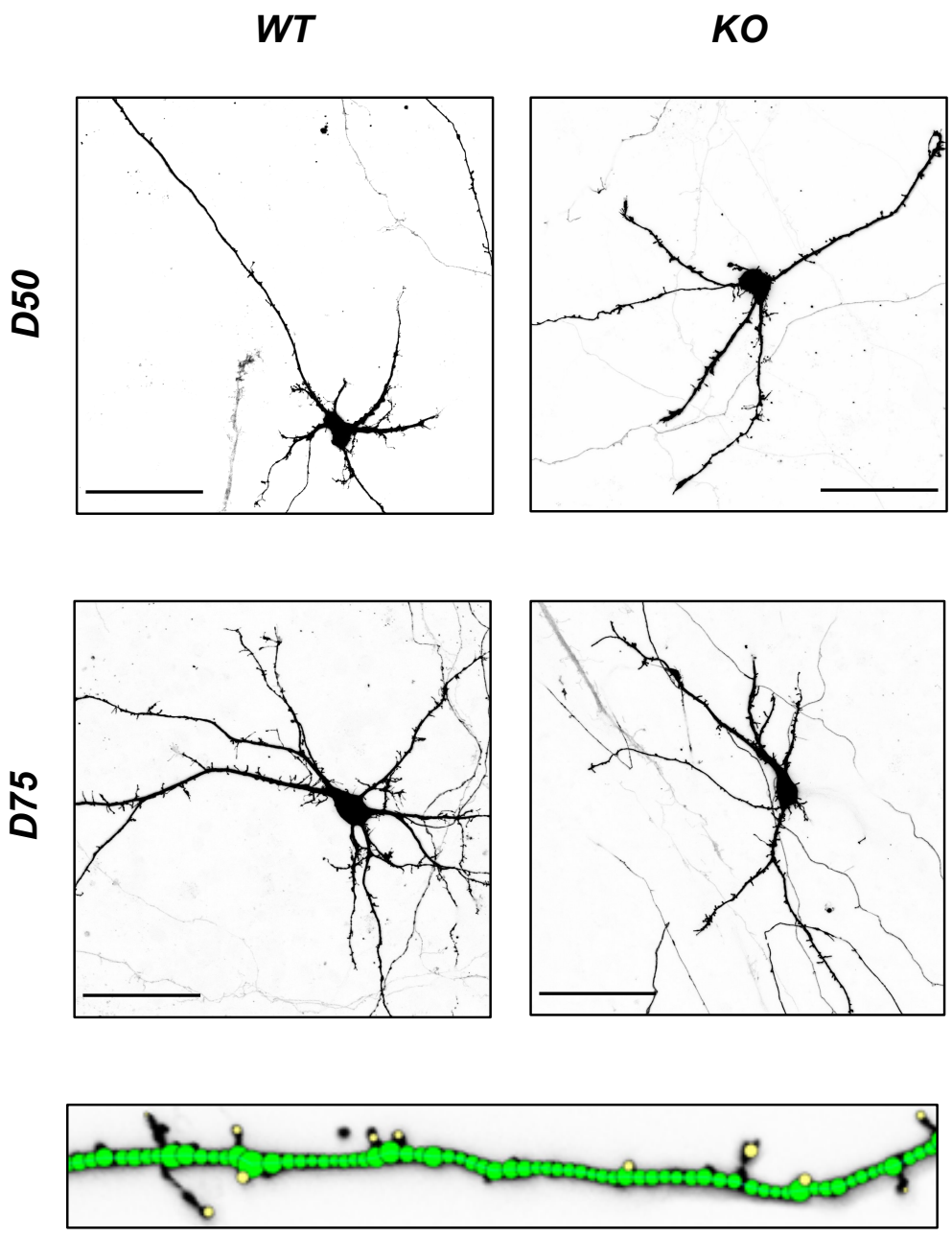


Figure 6.11. Example images of dendritic spine analyses

High resolution images were taken on an inverted confocal microscope (63X objective/1.35 numerical aperture). Neurons were sparsely labelled with mNeonGreen at D35 and plated separately by genotype. At either D50 or D75 the cells were fixed with 4% PFA/sucrose and imaged. Scale bar measures 50 µm. Neurite and dendritic spine measurements were acquired with NeuronStudio software. An example reconstruction is shown in the bottom frame overlaying the raw image. The branch trace is shown in green and the spines in yellow.

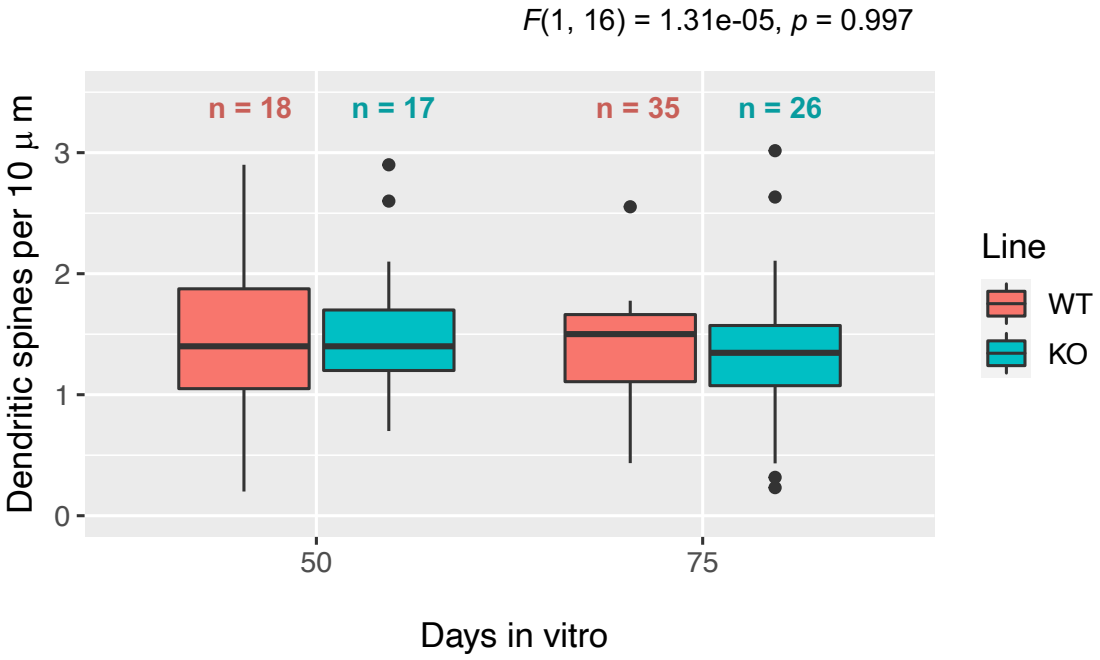


Figure 6.12. Average dendritic spine density per neuron

Neither *CNTNAP2* KO nor culture age had a significant effect on average dendritic spine density. KO results are shown above the graph; culture age results were calculated as $F(1, 16) = 3.00e\text{-}03$, $p = 0.958$. Post-hoc comparisons confirmed these findings, with $p > 0.05$ for all measurements.

6.4 Network-level neuronal activity

6.4.1 Results of the Incucyte neuronal activity assay

Given the evidence that *CNTNAP2* KO reduces neurite branching and total neuron length, my next question was to ask whether this potential decrease in connectivity is accompanied by changes in neural network function, given that the reduction in neuron length is likely to affect network architecture. It could be hypothesized that a reduction in neurite branching (and therefore, in neuron-to-neuron connectivity) could cause a likewise reduction in culture activity. In order to test this hypothesis, I set out to measure network-level activity in the WT and KO excitatory neuron cultures. This experiment relied on the use of **'neuroburst orange'**, a **calcium indicator lentivirus** driven by a Synapsin-1 promoter. Since synaptic activity in neurons involves the flux of **Ca^{2+} ions**, calcium indicators are routinely used to proxy for neuronal activity (382). Fluorescent indicators (including neuroburst orange), emit light upon binding to Ca^{+2}. Fluorescence therefore indicates an increase in intracellular, cytoplasmic Ca^{2+}, which in turn indicates a neuronal 'firing' event composed of one or more action potentials.

By measuring the number, intensity, and/or rate of fluorescent bursts, different aspects of culture activity can be quantified. **Table 6.01** describes the six parameters we considered: **1) number of active objects (cells/neurons)**, **2) mean object intensity**, **3) mean burst rate**, **4) mean burst duration**, **5) mean burst intensity**, and **6) mean burst synchronicity**. To measure these parameters, we used an Incucyte high-throughput cell imager. The Incucyte is a fluorescent microscope fitted into a 37°C / 5% CO_2 cell culture incubator. Every 24 hours, the microscope takes a **two-minute scan** of the cultures in its incubator; one scan per well of a 96-well imaging plate. Fluorescence emitted during this time is automatically translated into the six parameters by the Incucyte. Importantly, the scans were programmed to occur between 3 – 6 am every day, allowing for 'resting state' activity to be detected (i.e.

since handling/media changes, which could affect culture activity, would have occurred hours before).

Two inductions each of WT and KO cells were used, with each induction plated separately into 8 wells of a 96-well plate at 150,000 cells/cm^2. At D40, the cultures were infected with the neuroburst orange lentivirus, and then imaged every 24 hours from D41 until D106 (see **Figure 6.13**). Crucially, instead of using the usual **'N2B27'** neuron maintenance media, the base of which is DMEM/F12 media, we switched the cells to a media that is specially formulated to support neuronal activity (**'BrainPhys'**) (383). Previous experiments in the lab have shown almost no cell activity is detected with N2B27, but that robust activity can be observed with BrainPhys. Both WT and KO cells were transitioned to BrainPhys at D37, and kept in this media throughout the course of the experiment.

Culture activity was found to be highly correlated with media changes, with activity peaking ~12-24 hours after fresh BrainPhys was added. Peaks (local maximums) and troughs (all other data points) were therefore analyzed separately. A two-way repeated measures ANOVA was used to determine the effect of **1) culture age** and **2) *CNTNAP2* KO** on each of the six activity parameters. Multiple comparison tests were then used to assess WT versus KO at each timepoint (Sidak's test) and between timepoints within each line (Tukey test). Each of these analyses were performed for both the peak and trough datasets separately.

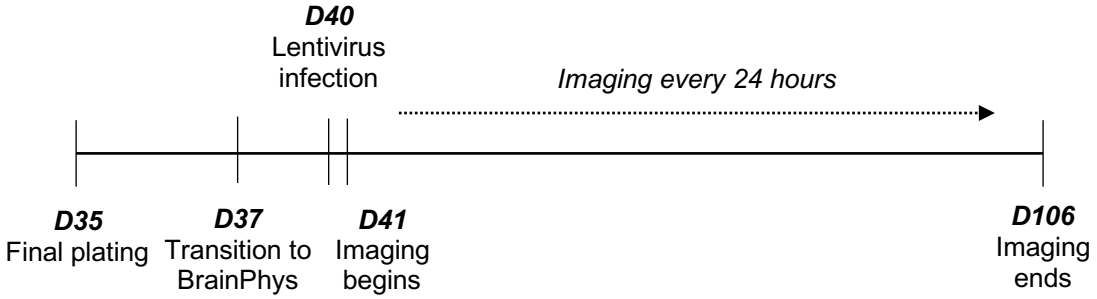


Figure 6.13. Culture activity experimental timeline

At D35, WT and KO cells were plated separately at 150,000 cells/cm^2 into a 96-well imaging plate. At D37 the cells were transitioned from N2B27 to BrainPhys media to support neuronal activity. At D40 the cells were infected with neuroburst orange lentivirus. Every 24 hours thereafter, the Incucyte took a two-minute scan of each well.

Parameter	*Description*
No. of active objects	No. of objects (cells or cell clusters) that burst at least once over the 2 min scan
Mean object intensity	Mean intensity of all objects over the 2 min scan
Mean burst rate	(Average number of bursts per 2 min scan)/(2 min)
Mean burst duration (sec)	Mean duration of bursts (in seconds)
Mean burst intensity	(Area under burst)/(burst duration)
Mean burst synchronicity	**Measure of network connectivity:** 0 = random 1 = highly synchronized

Table 6.01. Culture activity parameters

Descriptions of the six parameters measured by the Incucyte activity assay. Fluorescence emitted during the scans was measured by each parameter for both WT and KO neurons (plated separately).

Firstly, for the mean **number of active objects**, a two-way ANOVA detected significant effects of culture age and *CNTNAP2* KO in both the peak and trough datasets [Peak = **culture age:** $F(25, 364) = 295.6$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 364) = 3984$, $p < 0.0001$, and **interaction:** $F(25, 364) = 114.9$, $p < 0.0001$] [Trough = **culture age:** $F(39, 560) = 527.9$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 560) = 3257$, $p < 0.0001$, and **interaction:** $F(39, 560) = 297.4$, $p < 0.0001$]. Please refer to **Figure 6.14** for plots of the data.

Out of the 26 timepoints examined in the peak dataset and the 40 timepoints in the trough dataset, post-hoc comparisons found the majority were significantly different between WT and KO. Between ~D42 – D60, WT cultures had a greater number of active objects – largely in agreement with our hypothesis given the reduced branching observed in KO neurons between ~D50 – D60 [Peak = **D49:** $p < 0.0001$, and **D51:** $p < 0.0001$; Trough = **D45:** $p = 0.003$, **D46 – D59:** $p < 0.0001$]. From D61 onwards a much stronger difference emerges – this time with the KO cells showing several hundred more active objects than WT [Peak = **D61 – D100:** $p < 0.0001$ and **D106:** $p = 0.03$; Trough = **D63:** $p < 0.0001$, **D66 – D95:** $p < 0.0001$, **D101:** $p = 0.0007$, and **D104 – D105:** $p < 0.0001$]. Similarly, many pairwise comparisons between timepoints (within each line) were significantly different, indicating the number of active objects increased with culture age in both WT and KO cells. In general, active objects increased with age in both WT and KO lines until ~D70 when KO objects began to slowly decrease in number. WT objects on the other hand, appeared to be increasing at a relatively stable rate throughout the 60-day experiment.

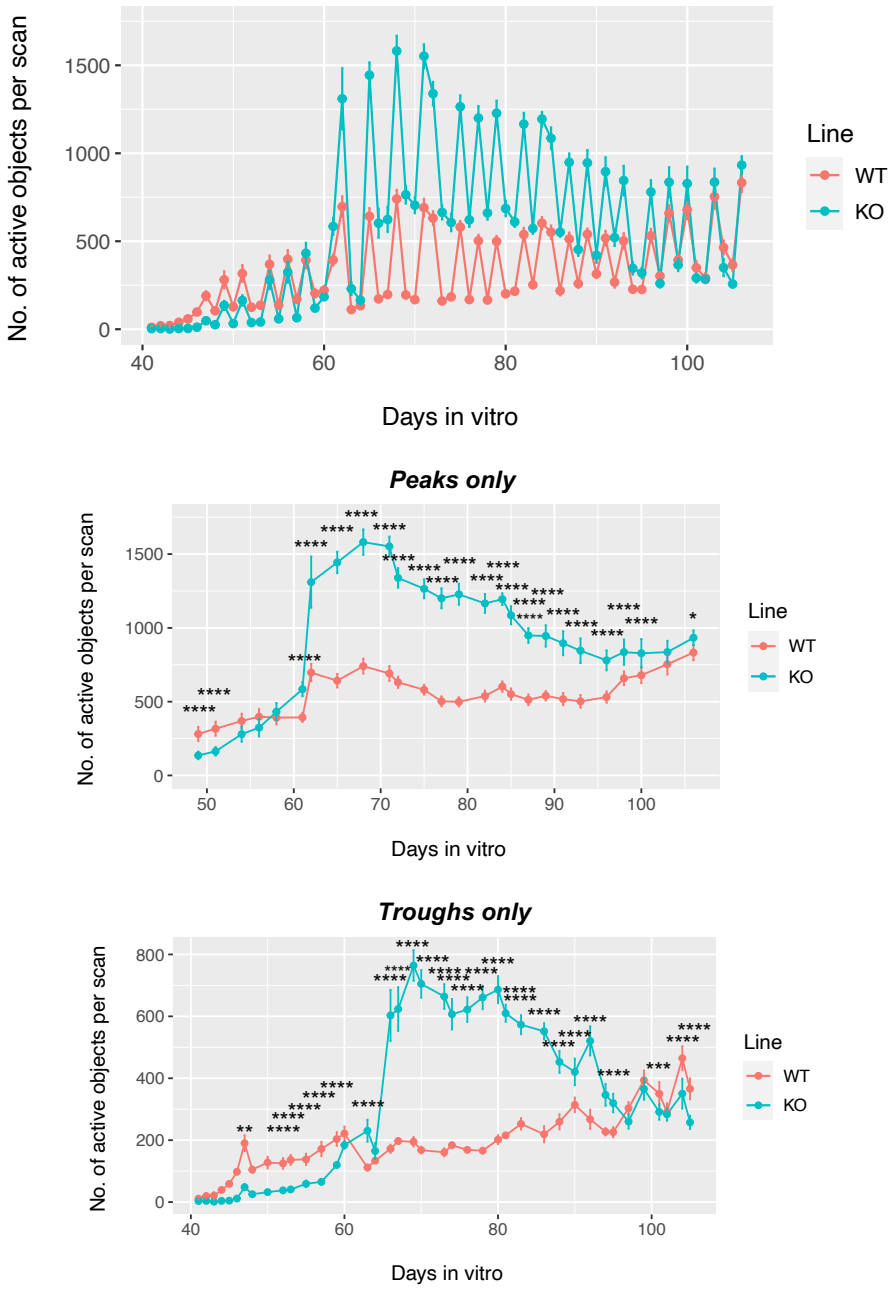


Figure 6.14. Number of active objects per two-minute scan

Active objects are defined as cells or cell clusters that fire at least once over a 2-minute scan. Culture activity was found to be highly correlated with media changes, with activity peaking ~12-24 hours after fresh BrainPhys was added. Peaks and troughs were therefore analyzed separately. For both datasets, two-way ANOVA found a significant interaction between the number of active objects with 1) culture age and 2) *CNTNAP2* KO. Error bars show standard error, with each data point representing the average of 8 wells per genotype. Significance asterisks are graphed for WT versus KO comparisons at a given timepoint. All other results are described in the main text. Overall, a clear increase in the number of active objects is observed in the KO cells.

Mean **object intensity** showed a very similar trend to the number of active objects (**Figure 6.15**). Both culture age and *CNTNAP2* KO were found to have significant effects on object intensity [Peak = **culture age:** $F(25, 364) = 311.1$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 364) = 1036$, $p < 0.0001$, and **interaction:** $F(25, 364) = 68.66$, $p < 0.0001$] [Trough = **culture age:** $F(39, 560) = 275.8$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 560) = 1060$, $p < 0.0001$, and **interaction:** $F(39, 560) = 103.3$, $p < 0.0001$]. In the trough dataset, object intensity was marginally higher in KO than WT cells between ~D50 – D60, while no significant difference was observed in the peak dataset during this time period [Trough = **D50:** $p < 0.01$, **D52 – D53:** $p < 0.0001$, **D55:** $p < 0.01$, and **D57 – D60:** $p < 0.0001$]. Again, around ~D61 KO cells begin to show a considerably larger object intensity, which persists until ~D90 when WT cells show slightly more intensity until D105 [Peak = KO higher at **D62:** $p < 0.0001$ and from **D68 – D93:** $p < 0.0001$, and lower at **D100:** $p < 0.0001$; Trough = KO higher at **D63:** $p < 0.0001$, **D66 – D88:** $p < 0.0001$, **D90:** $p < 0.001$, **D92:** $p < 0.0001$, and lower at **D97:** $p < 0.001$, **D99:** $p < 0.0001$, and **D102 – D105:** $p < 0.0001$]. In general, object intensity increased with age in both WT and KO lines until ~D70 when both lines began to slowly decrease.

Once again, significant effects by culture age and *CNTNAP2* KO were detected on mean **burst rate** by ANOVA [Peak = **culture age:** $F(25, 364) = 270.1$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 364) = 8750$, $p < 0.0001$, and **interaction:** $F(25, 364) = 100.8$, $p < 0.0001$] [Trough = **culture age:** $F(39, 560) = 155.3$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 560) = 869.6$, $p < 0.0001$, and **interaction:** $F(39, 560) = 68.21$, $p < 0.0001$] (see **Figure 6.16**). Sidak's multiple comparisons test reported KO cells had increased burst rates at the following timepoints: Peak = **D49:** $p < 0.0001$, **D51:** $p = 0.001$, **D56 – D58:** $p < 0.001$, and **D62 – D106:** $p < 0.0001$; Trough = **D63:** $p < 0.0001$, **D66 – D92:** $p < 0.0001$, **D95:** $p = 0.002$, **D101 – D102:** $p < 0.0001$, **D104:** $p = 0.009$, and **D105:** $p < 0.0001$. The trough dataset also reported WT had significantly increased burst rates early in the experiment (as with some of the other parameters) [**D41:** $p = 0.002$, **D46:** $p = 0.006$, **D48:** $p = 0.007$, **D52 – D53:** $p < 0.0001$, **D55:** $p = 0.0009$, and **D57 – D60:** $p < 0.0001$]. Both WT and KO cells show an increase in rate with culture maturity.

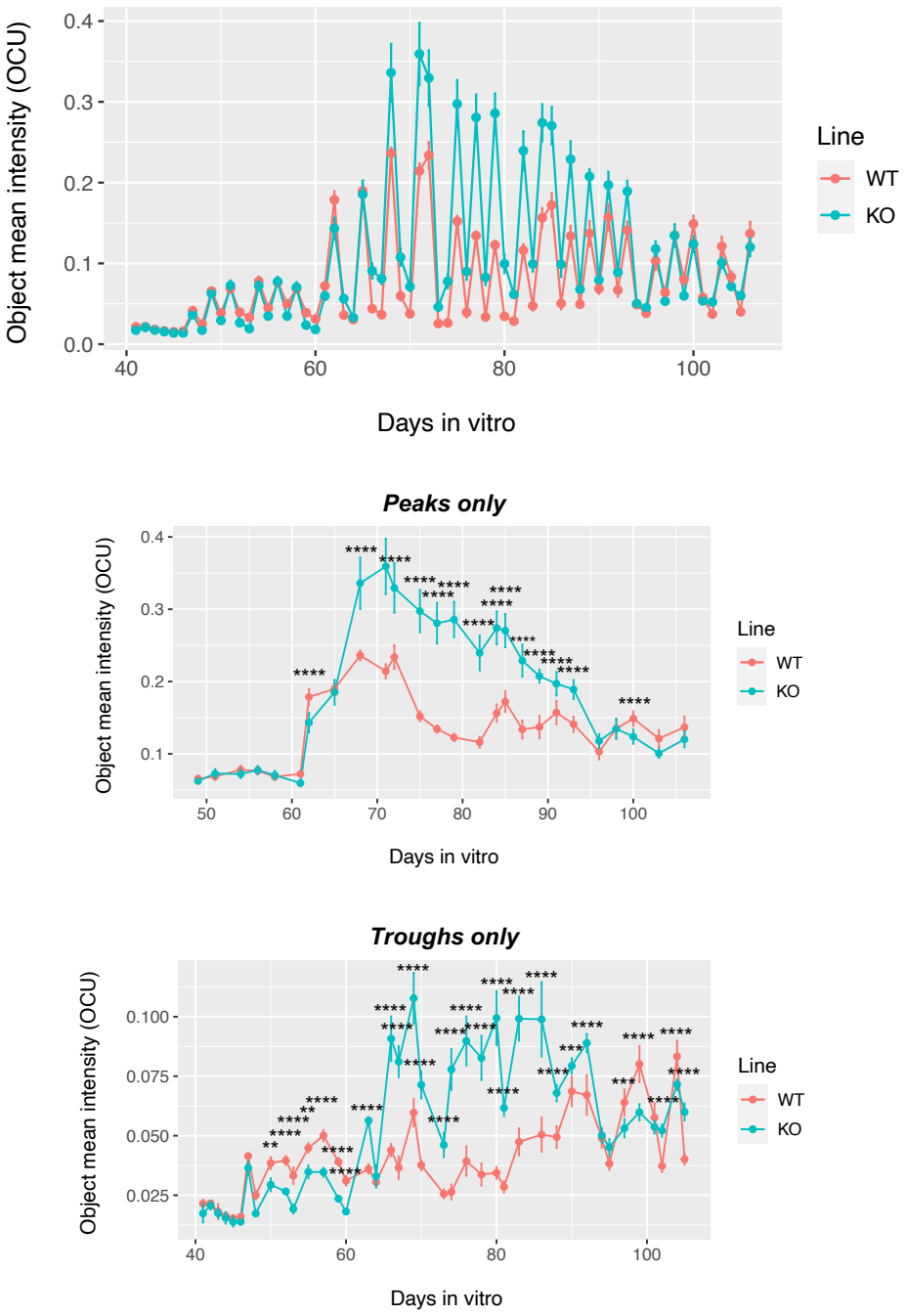


Figure 6.15. Mean object intensity per two-minute scan

For both peak and trough datasets, two-way ANOVA found a significant interaction between mean object intensity with 1) culture age and 2) *CNTNAP2* KO. Error bars show standard error, with each data point representing the average of 8 wells per genotype. Significance asterisks are graphed for WT versus KO comparisons at a given timepoint. All other results are described in the main text. Overall, a clear increase in object intensity is observed in the KO cells.

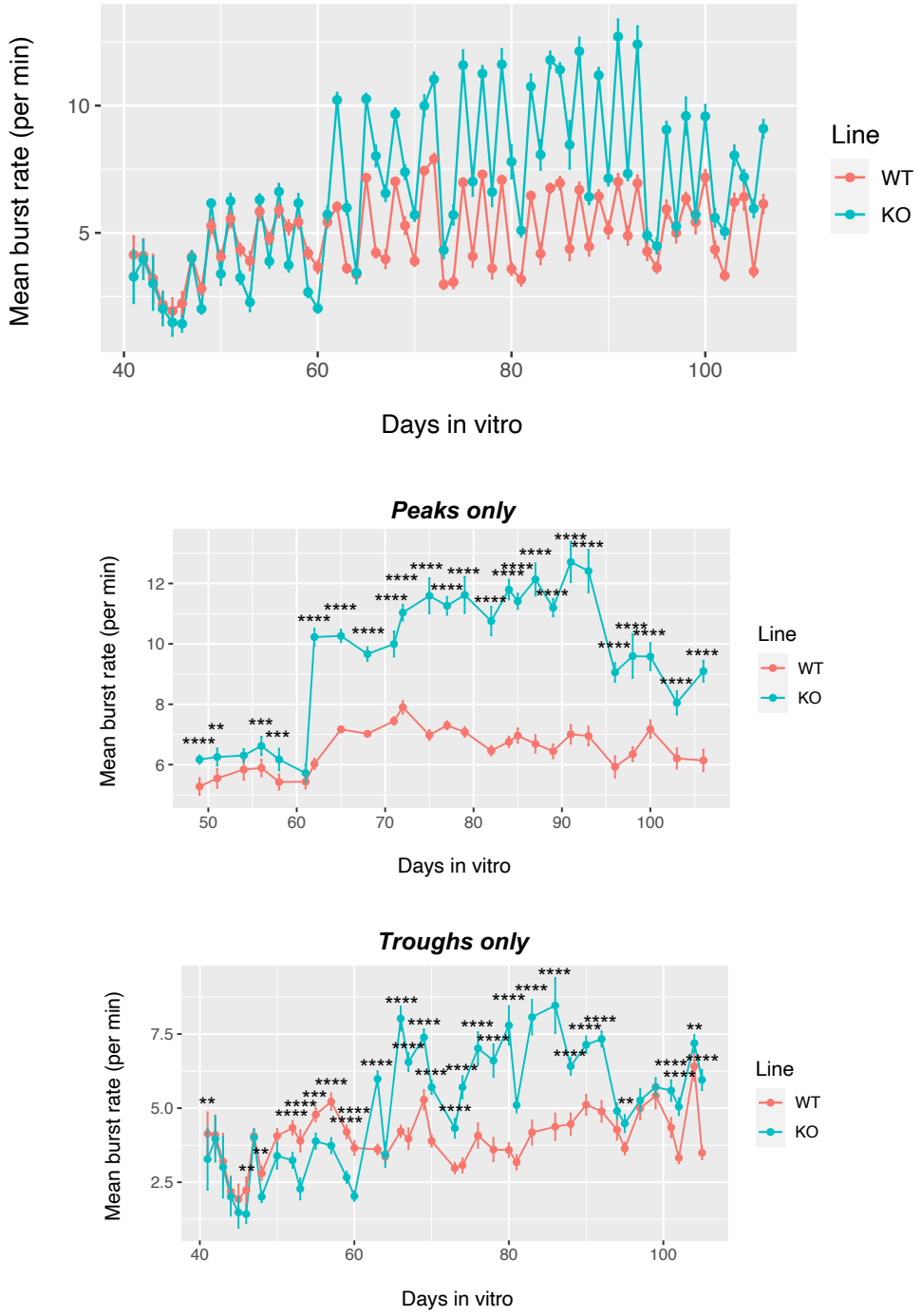


Figure 6.16. Mean burst rate per two-minute scan

For both peak and trough datasets, two-way ANOVA found a significant interaction between mean burst rate with 1) culture age and 2) *CNTNAP2* KO. Error bars show standard error, with each data point representing the average of 8 wells per genotype. Significance asterisks are graphed for WT versus KO comparisons at a given timepoint. All other results are described in the main text. As with the other parameters examined, a clear increase in burst rate is observed in the KO cells.

For mean **burst duration**, once more a two-way ANOVA revealed both culture age and *CNTNAP2* KO had a significant effect [Peak = **culture age:** $F(25, 364) = 294$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 364) = 4771$, $p < 0.0001$, and **interaction:** $F(25, 364) = 83.31$, $p < 0.0001$] [Trough = **culture age:** $F(39, 560) = 242$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 560) = 7.923$, $p = 0.005$, and **interaction:** $F(39, 560) = 111.2$, $p < 0.0001$] (see **Figure 6.17**). Notably, for this parameter WT cells were significantly increased relative to KO at the majority of timepoints examined. In the Peak dataset increases in burst duration were noted in WT neurons at **D54:** $p < 0.0001$, **D61 – D75:** $p < 0.0001$, **D77:** $p = 0.0002$, **D79 – D82:** $p < 0.01$, and **D84 – D106:** $p < 0.0001$. In the Trough dataset WT duration was higher than KO at the beginning and end of the experiment [**D41 – D42:** $p < 0.0001$, **D45:** $p = 0.003$, **D47 – D60:** $p < 0.0001$, and then again at **D90 – D92:** $p < 0.0001$, **D97 – D101:** $p < 0.0001$, **D104:** $p < 0.0001$, and **D105:** $p = 0.015$]. From ~D63 – D83 KO cells showed significantly longer burst durations [**D63:** $p < 0.0001$, and then **D66 – D83:** $p < 0.0001$].

Similar to object intensity, mean **burst intensity** was also significantly up-regulated in KO versus WT. ANOVA once more found culture age and *CNTNAP2* KO had a significant effect on burst intensity [Peak = **culture age:** $F(25, 364) = 247.3$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 364) = 591.6$, $p < 0.0001$, and **interaction:** $F(25, 364) = 77.2$, $p < 0.0001$] [Trough = **culture age:** $F(39, 560) = 471.5$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 560) = 929.6$, $p < 0.0001$, and **interaction:** $F(39, 560) = 137.3$, $p < 0.0001$] (see **Figure 6.18**). In the Peak dataset burst intensity was significantly increased in KO cells at **D68 – D89:** $p < 0.0001$ and **D93:** $p = 0.0162$, and decreased in KO cells at **D62:** $p < 0.0001$, **D98:** $p = 0.006$, and **D100 – D106:** $p < 0.0001$. Similarly, in the Trough dataset intensity was higher in KO between ~D65 – D90 [**D63:** $p < 0.0001$, **D66 – D88:** $p < 0.0001$, and **D90:** $p = 0.0007$]. At earlier (~D40 – D60) and later timepoints (D90 – D105) WT cells had higher burst intensity [**D41:** $p = 0.0214$, **D47:** $p = 0.0009$, **D48 – D50:** $p < 0.01$, **D52 – D53:** $p < 0.0001$, **D55:** $p = 0.0022$, **D57:** $p = 0.0007$, **D59 – D60:** $p < 0.0001$, and again at **D92:** $p < 0.0001$, **D97 – 101:** $p < 0.0001$, and **D104:** $p < 0.0001$]. Overall, intensity in KO cells increased until ~D70, after which it began to

slowly decrease. In WT cells, intensity generally increased throughout the experiment (with some appreciable fluctuation around the middle timepoints).

Finally, for mean **synchronicity**, once more both culture age and *CNTNAP2* KO were detected to have a significant effect [Peak = **culture age:** $F(25, 364) = 342.4$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 364) = 1435$, $p < 0.0001$, and **interaction:** $F(25, 364) = 108.8$, $p < 0.0001$] [Trough = **culture age:** $F(39, 560) = 259.7$, $p < 0.0001$, ***CNTNAP2* KO:** $F(1, 560) = 1607$, $p < 0.0001$, and **interaction:** $F(39, 560) = 203.6$, $p < 0.0001$] (see **Figure 6.19**). In the Peak dataset KO neurons showed higher correlation in their activity between D62 – D96 [**D62 – D89:** $p < 0.0001$ and **D91 – D93:** $p < 0.01$]. At the final few timepoints WT were more highly synchronous [**D96:** $p < 0.0001$ and **D100 – D106:** $p < 0.0001$]. Similarly, in the Trough dataset KO bursts were significantly more correlated between D66 – D95 [**D66 – D94:** $p < 0.0001$ and **D95:** $p < 0.001$]. Between D96 – D105, KO synchronicity declined steeply while WT continued to rise. As such, WT neurons were significantly more synchronous during this period [**D96 – D105:** $p < 0.0001$]. In both peak and trough datasets KO synchronicity increased steeply from ~D60 – D90, after which synchronicity declined rapidly. In WT cells conversely, synchronicity increased gradually throughout the 60-day experiment.

Therefore, *CNTNAP2* KO neurons appeared to be overall more active than WT cells based on the six parameters examined. Specifically, KO neurons had a greater number of active objects that were more intense. They also showed faster, stronger, and more highly synchronized network bursts. The implications of these findings will be discussed further in **Section 6.5**.

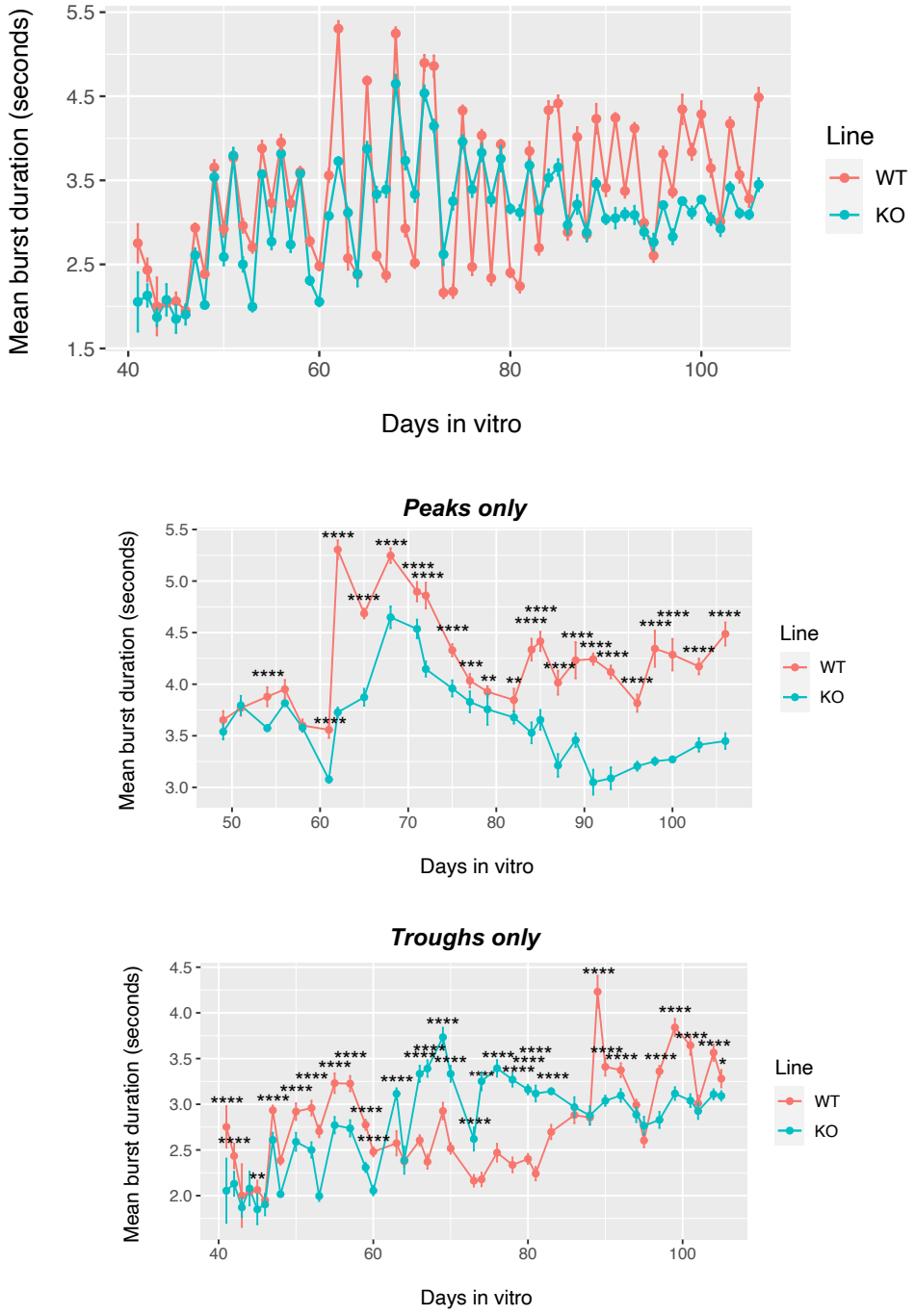


Figure 6.17. Mean burst duration per two-minute scan

For both peak and trough datasets, two-way ANOVA found a significant interaction between mean burst duration with 1) culture age and 2) *CNTNAP2* KO. Error bars show standard error, with each data point representing the average of 8 wells per genotype. Significance asterisks are graphed for WT versus KO comparisons at a given timepoint. All other results are described in the main text. Unlike the other parameters examined, an increase in burst duration is observed in WT cells in the peak dataset, and a less clear trend in the trough dataset.

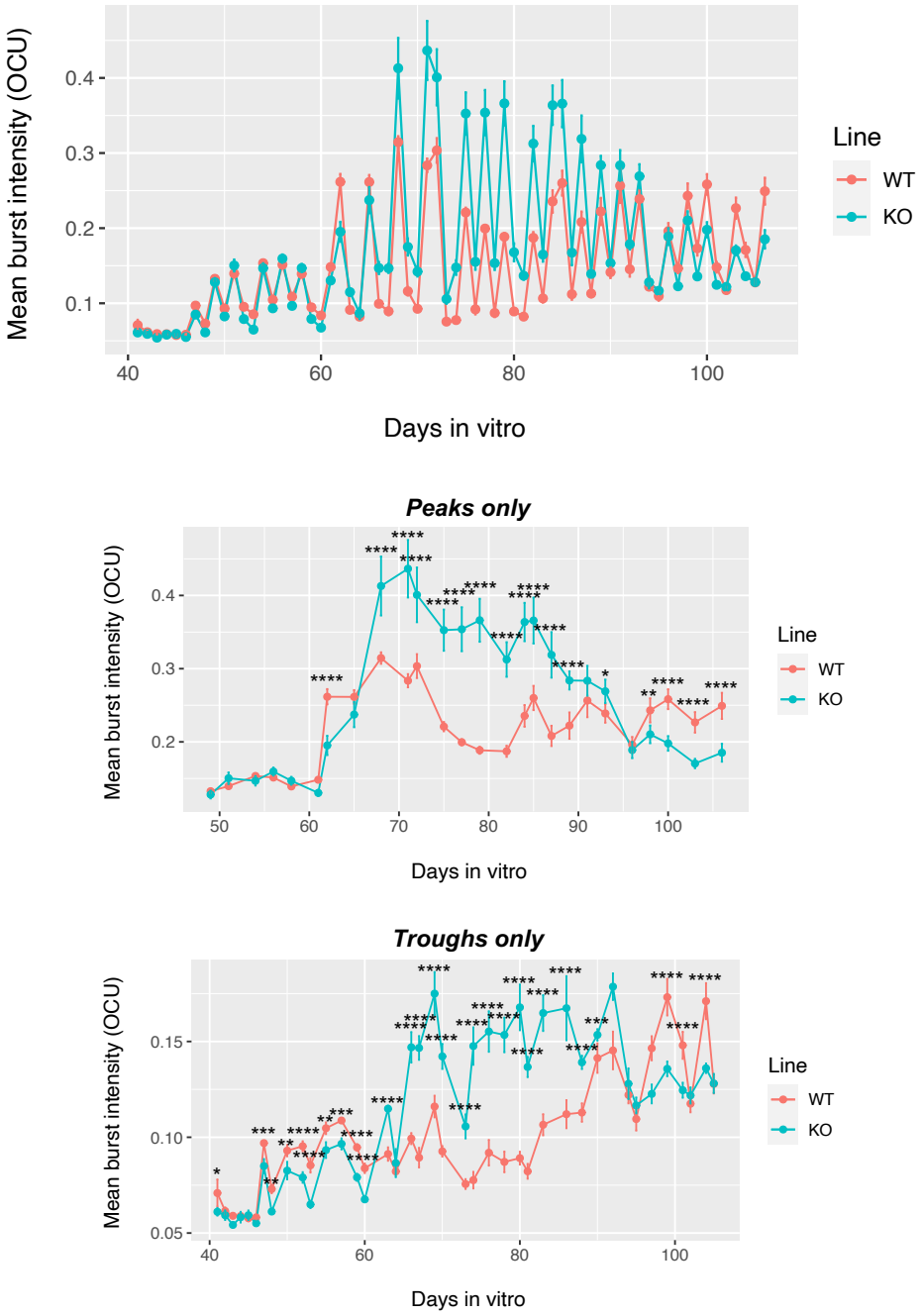


Figure 6.18. Mean burst intensity per two-minute scan

For both peak and trough datasets, two-way ANOVA found a significant interaction between mean burst intensity with 1) culture age and 2) *CNTNAP2* KO. Error bars show standard error, with each data point representing the average of 8 wells per genotype. Significance asterisks are graphed for WT versus KO comparisons at a given timepoint. All other results are described in the main text. As with most of the other parameters examined, a clear increase in burst intensity is observed in the KO cells.

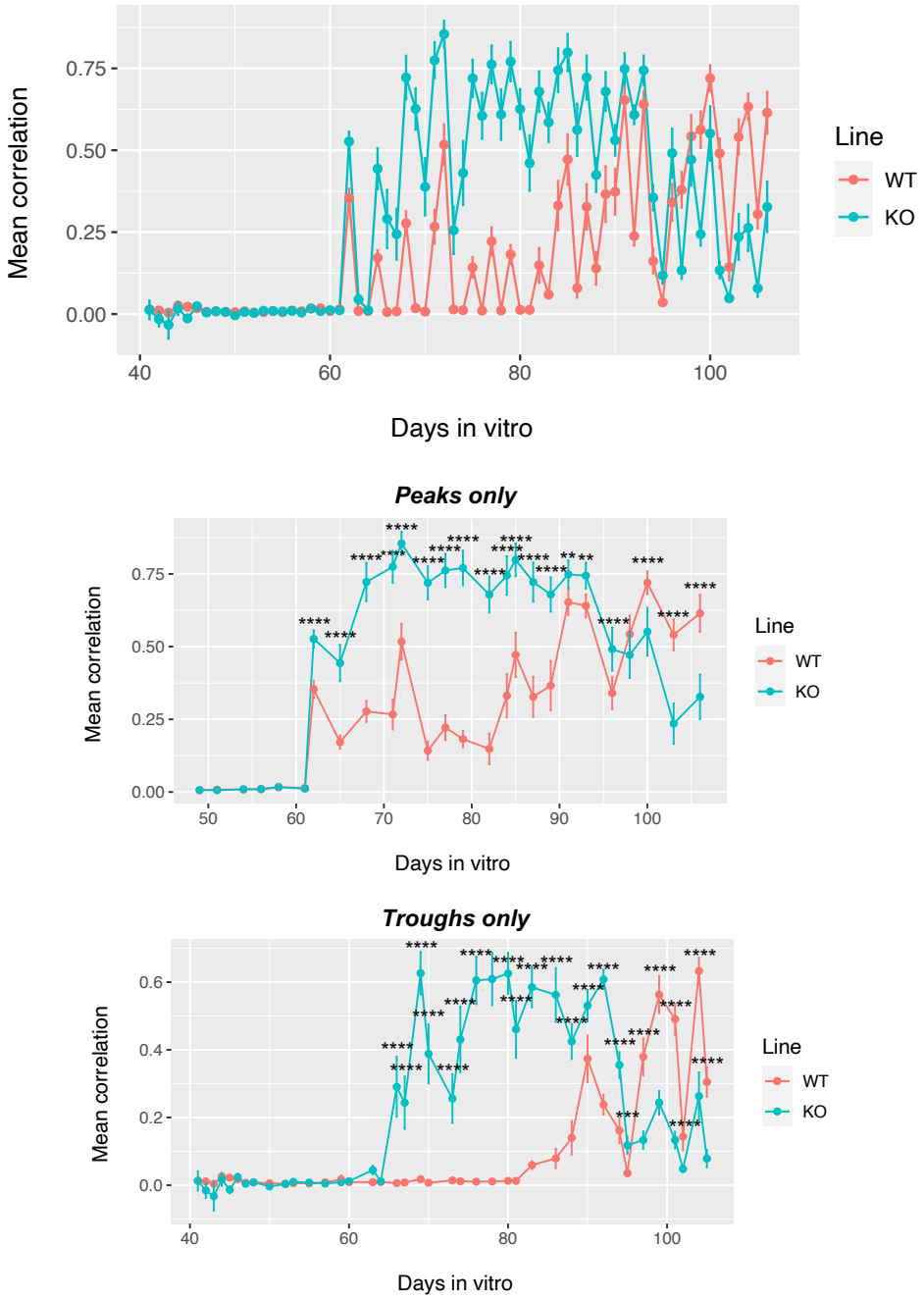


Figure 6.19. Mean burst synchronicity per two-minute scan

For both peak and trough datasets, two-way ANOVA found a significant interaction between mean burst synchronicity with 1) culture age and 2) *CNTNAP2* KO. Error bars show standard error, with each data point representing the average of 8 wells per genotype. Significance asterisks are graphed for WT versus KO comparisons at a given timepoint. All other results are described in the main text. A clear increase in burst synchronicity is observed in the KO cells.

6.5 Discussion

In this chapter, I report on a number of complementary experiments that addressed the function of *CNTNAP2* in human forebrain development. Firstly, I examined what effect loss of *CNTNAP2* has on 1) neurite length, 2) total neuron length, and 3) the number of branches per neuron. Plating WT and KO cells **separately** revealed no changes in **neurite length** or **total neuron length** between the two genotypes (**Figures 6.03-6.04**). **Branch number** was only significantly different at one timepoint (D55: reduced branching in KO cells, $p = 0.03$) (**Figure 6.05**). Plating WT and KO cells in the **same wells** revealed somewhat different results. **Branch length** was detected as significantly higher in WT than KO at D48 and D60 ($p = 0.0171$ and $p = 0.006$, respectively; see **Figure 6.08**). However, this trend was not consistent, with KO cells frequently having comparable (or even slightly higher) branch lengths at other timepoints. Taken together with the findings of the first experiment, the evidence for an effect of *CNTNAP2* KO on branch length is weak. It also suggests that the reduced branch length observed in *Cntnap2* KO mouse models (35, 202, 261) may not occur in humans. This is, of course, subject to further experimentation to confirm whether these findings are replicated or not.

Perhaps more interesting, were the findings that KO cells had reductions in **total length** and **branch number** in the co-cultures (**Figures 6.09-6.10**). Unlike for neurite length, these observations fall in line with previous data from mouse studies. As stated, cortical cultures from *Cntnap2* KO/KD mice have shown reduced dendritic branching (35, 202). Resting state functional MRIs (rsfMRIs) have independently detected *Cntnap2* null mice have significantly reduced local and long-range cortical connectivity (264, 265). It also follows from logic that a decrease in branching should correlate with a decrease in total length.

These results fit with a potential role for *CNTNAP2* in diseases linked with aberrant cortical connectivity. Whether reduced dendritic branching results in a loss

of connectivity would need to be definitively shown (i.e. through pseudo-rabies tracing). However, in **Chapter 2**, I outlined some general trends in connectivity associated with ASD and schizophrenia. Although autism has been connected to a trend of hyper-connectivity, there are many known exceptions to this generalization (384-386). As such, it is not beyond the scope of the expected to see reduced neurite branching. Fewer branches could mean fewer neuron-to-neuron connections are made (hypo-connectivity). With respect to schizophrenia, these results do fit with a disease mechanism of reduced connectivity (94). However, not a single case of schizophrenia has been reported in an individual with homozygous loss of *CNTNAP2*. These disorders were limited to ASD, Pitt-Hopkins syndrome, and cortical dysplasia focal epilepsy (189). Therefore, there is little evidence to propose a link between this phenotype and schizophrenia (although one could be possible).

The finding that *CNTNAP2* KO had an effect on total length/branching in the co-cultures, but no effect when WT and KO were plated separately, could be explained by one of two reasons. Firstly, it is possible that when the WT and KO neurons were plated separately, differences in culture composition developed over time. The KO cells could have grown to a slightly higher density, and this could have promoted an increase in neurite branching. As mentioned, cell density is known to stimulate neurite and dendritic spine growth (387). A 'true' difference in branching could therefore have been obscured - and only detectable once the two genotypes were plated together. This has been suggested previously by other groups studying neurite growth between WT and KO cell lines (380). Additionally, given the increase in sample size between the first and second experiment, it could simply be that the first experiment (unmixed) did not have the power to detect the changes.

As discussed in **Section 6.2.3**, the measurements taken in experiment #2 (co-cultures) were significantly larger than the measurements taken in experiment #1 (no mixing) at certain timepoints. One explanation could be that because the experiments were performed separately, using different batches of reagents, that the nucleofection efficiency in the earlier experiment was not as high as in the second. If the second

experiment had, for example, more fluorescent protein incorporated into the cells, this could fill the neurites better and explain the larger readings. My own improvements in nucleofection technique between the two experiments could also explain this. Although both sets of experiments were final plated at the same density (150,000 cell/cm^2), later differences in cell density could also be at play (as proposed for why no change in branching was originally observed). A greater number of neurons, by chance, in the second experiment could have promoted greater neurite outgrowth/branching (387).

Another curious feature of the neurite length/branching assay was the noticeable decrease in all three parameters after ~D55. Given that neurites would be expected to lengthen with development, and considering 1) the age of the cultures and 2) the diminishing image quality, I suspect that fluorophore bleaching was occurring from the frequent imaging. The young age of the cultures means it would be too early for neurite pruning to be a plausible explanation (1). Subsequent follow-up studies should investigate this, by starting imaging at later timepoints and/or imaging less frequently (i.e. preventing bleaching until later in culture development).

Moreover, although the **dendritic spine assay** was only a small experiment, its results still warrant discussion. As shown in **Figure 6.12**, no difference in spine density was observed between WT and KO at either D50 or D75. There was also no difference in density between timepoints/within the genotypes. However, this experiment should be repeated with a larger sample size before more definitive conclusions can be drawn. They should also be conducted at later timepoints. D75 may simply be still too early to see an effect. Instead of imaging fixed neurons, super resolution microscopy could also be used to examine spine dynamics. This would be particularly useful given the finding by Gdalyahu et al. (36) that *Cntnap2* KO mice had decreased stability of newly formed spines. No reduction in the formation of new spines was observed, nor was any effect on the maintenance or pruning of already-formed/stable spines. Consequently, it may be that the average density of spines is the same between the two genotypes, but that the rate of spine turnover is different. This could be highly

informative given that the generation, pruning, and shape-changing of spines are essential for short- and long-term brain plasticity (122). It would also be worthwhile measuring not only spine density, but synapse density (the ultimate readout). This can be performed relatively simply, by staining for pre-synaptic and post-synaptic proteins and measuring their overlap using ImageJ/Fiji plug-ins.

The final section of this chapter focused on changes to the network-level activity of *CNTNAP2* KO and WT cultures. A striking pattern emerged in these experiments: starting from ~D60, the KO cultures became much more active than their WT counterparts. Activity was measured by 1) number of active objects, 2) mean object intensity, 3) mean burst rate, 4) mean burst duration, 5) mean burst intensity, and 6) mean burst synchronicity (**Figures 6.14-6.19**).

These data are extremely interesting in light of the connection between neuronal over-excitation and autism (189). It could also explain the frequent seizures these patients experience. Notably, this phenotype has been previously described in the literature. Flaherty et al. (239) studied PSC-derived neural progenitor cells from a schizophrenic patient with a heterozygous *CNTNAP2* mutation. The authors detected an over 200% increase in spontaneous network level synaptic activity, as well as abnormal expression of several synapse genes. It does not, however, agree with the majority of the literature on *Cntnap2* KO mouse models. For example, two studies showed KO/KD of mouse *Cntnap2* decreased the amplitude and frequency of excitatory and inhibitory mini post-synaptic currents (35, 263). One of the studies (263) also reported KO neurons had reduced network synchrony – the opposite to what was observed in our data. These discrepancies could relate to human versus mouse differences, the lack of interneurons in our cultures, or reflect homozygous versus heterozygous loss of the gene.

A separate question to consider is why our *CNTNAP2* KO cultures were detected as having reduced branching but were more active. As mentioned at the start of the chapter, my hypothesis was that greater connectivity would translate into

greater activity (380). This discrepancy could be explained by a loss of inhibitory neurons. If the KO cells are receiving less inhibition, this could cause an excitation : inhibition imbalance (i.e. overactivity). Loss of cortical interneurons was previously reported in a number of *CNTNAP2* KO mouse studies (199, 245, 271, 272). That said, our scRNA-Seq data from the previous chapter showed that there were no cortical interneurons in either the WT or KO cultures. Therefore, there cannot be a difference in their excitation : inhibition ratios.

Another explanation relates to the fact that our neurite length/branch assay ended at D66. A number of the Incucyte parameters, however, were only beginning to be increased in KO cells around this time. Mean burst intensity, mean burst duration, mean burst rate, and mean object intensity all showed WT cells had higher values until ~D60. Therefore, it's possible that the WT cells were both better connected and more active initially, and that the KO cells showed greater connectivity after the neurite length/branching assay ended. In other words, there could be a delayed development of neuronal connectivity/activity in *CNTNAP2* KO cells, but they later become over-connected/over-active. Repeating the neurite length/branching assay at later timepoints will be essential to clarifying if this is indeed the case.

A final point to discuss was the decrease in activity towards the end of the Incucyte experiment. From about D70 onwards, the measurements for almost all of the parameters decreased. The difference between WT and KO also diminished. The strongest reason for why this could have occurred relates to the stress placed on older *in vitro* cultures. By these timepoints, many cells begin to detach from the culture plates. This decrease in activity could therefore be caused by dying cultures. The fact that culture density increases with age could be an additional reason. Higher density cultures deplete nutrients in the culture media much more rapidly. They also accumulate metabolic waste at a similarly increased rate. The decrease in activity could therefore reflect that the cultures were being metabolically stressed.

A few important next steps will be needed to fully clarify the interpretation of this chapter's results. As always, further experiments with more cells and cell types will be needed (particularly for the Incucyte assay, which was only run once). One could also replicate the experiments using different techniques (e.g. multi-electrode assay (MEA) or patch-clamping instead of using the Incucyte). What these initial results do suggest, however, is that loss of *CNTNAP2* in human cortical cultures appears to cause a significant perturbation to neuronal connectivity/activity.

These results are highly useful for understanding the phenotypes caused by *CNTNAP2* mutations (e.g. seizures, autism), as well as highlighting possible therapeutic avenues. With respect to evolution, it is slightly more complicated. A *CNTNAP2* KO line has complete loss of the mRNA/protein – however, non-human primates still express *CNTNAP2*, just less strongly[8]. It is not clear whether a reduction in *CNTNAP2* would have the same effects on connectivity/activity as a complete loss. Considering that heterozygous mutations also cause disorders of synaptic function, it does still suggest even partial reductions in *CNTNAP2* have important cognitive consequences. In light of this, these results offer further support that *CNTNAP2* could be involved in cognitive evolution, through an increase in human cortical expression, leading to some alteration in human synaptic function. While electrophysiological comparisons of human and primate neural activity are rare, there is some initial data to suggest there are significant differences in cortical activity between humans and our primate relatives (29, 388).

[8] *If the finding of reduced *CNTNAP2* expression in non-human primates (by us and others) is indeed correct.

Chapter 7

Evolutionary and functional studies of the *CNTNAP2* HARs

7.1 Introduction

As discussed in **Chapter 2**, many human accelerated regions are predicted to be gene regulatory elements (2, 170, 171, 175, 389). Capra et al. (170) used existing functional genomics data, in combination with machine learning algorithms, to show that 60% of non-coding HARs overlap epigenetic enhancer marks. Half of these were predicted to target genes active during development, and one third to act in the brain. Where the field is currently lacking, is backing up these predictions with hypothesis-driven experiments. While some progress has been made testing individual HARs for enhancer potential (e.g. using transgenic HAR::LacZ mice (171, 172, 389)), the vast majority of human accelerated regions have not been functionally characterized.

Of the HAR experiments that do exist, very few have been conducted in human tissue (most have been in mouse or zebrafish (2, 172, 390)). Given HARs are hypothesized to be human-specific enhancers, testing their enhancer potential in a human system could be immensely important for capturing their true function(s). Human PSC-derived models provide a useful means for doing just this. HAR function could be tested in a human biological context and temporally across cell development. Since enhancers usually act within a tightly controlled timeframe (391), this system allows such effects to be easily picked up during experimentation.

With respect to the *CNTNAP2* HARs, previous studies have highlighted the potential of several to act as human-specific enhancers. HACNS_884 was shown by Gittelman et al. (275) to overlap a human-specific DNase I hypersensitive site (DHS). Won et al. (175) further identified six of the eight HARs as overlapping DHSs in fetal brain (all except 2xHAR.395 and ANC1208). Finally, Capra et al. (170) detected HACNS_884 and HACNS_954 as putative enhancers using their enhancer finding pipeline (discussed in **Chapter 1**). They were also able to bioinformatically predict that HACNS_884 is active in fetal brain, but could not provide a prediction for HACNS_954. No studies have tested any of the eight HARs in an experimental enhancer assay. Given the interest of these sequences for human evolutionary and medical genetics, doing so could provide important information about the emergence of human cognitive abilities. As many of the *CNTNAP2* HARs also overlap disease-causing mutations (see **Figure 2.09**), understanding their function(s) could also clarify the causes and potential treatments of these illnesses.

As mentioned in **Chapters 1** and **2**, a minority of HARs are believed to be caused by GC-biased gene conversion (gBGC) or by a relaxation of constraint (i.e. and not positive selection) (169). Illuminating which of these mechanisms caused the *CNTNAP2* HARs to be accelerated will be important to understanding their significance. If the HARs are functional, it could also clarify their mechanism of action.

For example, the human accelerated region, HAR2, became a human-specific enhancer through gBGC-mediated loss of its repressor activity (9, 12).

Notably, there are existing reports of positive selection at the human *CNTNAP2* locus (see **Chapter 2**). Ayub et al. (40) identified selective sweeps in introns 1 and 13 in North-west European, East Asian, and Nigerian populations. These signals were detected by a combination of neutrality tests, including **Tajima's D** and **Fay and Wu's H**. Interestingly, a second study also identified signatures of selection in the same two introns (and the same three populations) (41). This time, however, the authors used the **F_{ST} statistic** - offering independent corroboration of the Ayub et al. findings. This second paper also applied the **K_a/K_s test** to look for positively selected protein-coding changes in *CNTNAP2*. While they did not find evidence of such selection on the human lineage, they did identify positive selection on the branch leading to bats *(Chiroptera)*. Since echolocation in bats is a vocal learning process (392), and *CNTNAP2* has been implicated in specific language impairments (195), this is still a highly interesting result.

Detailed neutrality tests have never been carried out at the *CNTNAP2* HARs specifically. The existing selection studies were also limited to only three populations (North-west European, East Asian, and Nigerian). More work is needed to clarify whether positive selection has occurred at these loci, and in what human populations. Under the assumption that *Homo sapiens*-specific traits should be shared by all human populations, any signature of selection operating at the species-level should be observed in all continental populations. A signature could also be interesting if it is found in African populations (whose genomes were not affected by the Out-of-Africa bottlenecks) (63). While population-specific signatures could still be noteworthy, they are more likely to represent intra-human differences, and not human-primate differences.

In this chapter, I will test for signatures of natural selection at the eight *CNTNAP2* HARs. I will then bioinformatically analyze the HARs for epigenetic enhancer marks, followed by a luciferase assay to experimentally test the HARs for enhancer function. Finally, I will contextualize our findings and discuss their implications for human evolution and disease.

7.2 Introduction to neutrality tests

Natural selection can be categorized into three main types: **1) positive selection, 2) negative selection,** and **3) balancing selection**. Positive selection causes beneficial traits - those that allow increased survival and reproduction - to become more frequent in the gene pool over time (63). Conversely, negative selection (also called purifying selection) drives deleterious traits to removed. Since new mutations are more likely to be harmful than helpful, most genetic changes will be subjected to purifying selection. Thirdly, **balancing selection** maintains the presence of multiple alleles in a population. This can occur through **heterozygote advantage** or **frequency-dependent selection**. A classic example of heterozygote advantage occurs at the hemoglobin-B *(HBB)* locus (393). Heterozygotes are both protected against severe sickle cell anemia (through having one non-sickle allele) and against malaria (by having one sickle allele).

Most selection tests have focused on positive selection, not only because it is the main driver of adaptation, but because it leaves clear signatures in the DNA. According to the **neutral theory of evolution**, most of the genetic variation within (and between) species is functionally neutral (167). These variants have not been selected for by natural selection, but have remained in the population by chance (a phenomenon called **genetic drift**). Tests for natural selection (also called **neutrality tests**) work through comparing the genetic variation at a locus with the expected variation under neutrality.

In this section, I will perform tests for positive and balancing selection on the *CNTNAP2* locus. Before doing so, however, I will briefly introduce the five main signatures of selection that are commonly tested for (adapted from Sabeti et al. (168), and summarized in **Table 7.01**). Importantly, caution must be taken when interpreting the findings of all selection tests. The demographic history of a population - including bottleneck events (a severe reduction in population size), population expansions,

and/or population sub-divisions - can also leave the same signatures as natural selection. Therefore, a positive result from a selection scan does not necessarily mean selection has occurred. Demographic explanations will be considered for all tests, along with the possibility of true selection.

7.3 Signatures of positive selection

7.3.1 Excess of non-synonymous substitutions (timescale: several MYA)

This signature of (positive) **directional selection** assumes that most **non-synonymous** substitutions are deleterious, while most **synonymous** substitutions are functionally neutral. Non-synonymous changes that are common or fixed (i.e. at 100% frequency) will therefore be rare in the genome. By this same logic most of the non-synonymous substitutions that have fixed are likely to be beneficial – and positively selected for. Therefore, an excess of non-synonymous substitutions at a locus is a potential signature of directional positive selection. Crucially, for a variant to reach high frequencies in a population, selection must have been acting over a significant period of time (394). This signature is therefore a marker of selection that occurred millions of years ago.

A commonly used neutrality test (called the **K_a/K_s test** (394)) compares the ratio of **non-synonymous to synonymous** substitutions with the identical ratio in a closely related species. For example, chimpanzees – who split from humans 5 – 13 MYA (45) – can be used as a reference for the ratio in 'pre-selection state'. Unfortunately, tests such as this can only be used on protein-coding DNA, and not intronic or intergenic regions. In non-coding DNA the lineage-specific acceleration of nucleotide substitutions (i.e. HARs for humans) has been used to infer positive selection instead (11). Inferring selection from this signature typically also requires multiple positively selected changes before a gene registers against background levels of neutral changes. In other words, positively selected individual changes could be missed.

7.3.2 Reduction in genetic diversity (timescale: < 250 KYA)

When a positively selected variant increases in frequency, nearby linked variants also increase in frequency – a process referred to as a **selective sweep** (63). Sweeps cause genetic diversity in the region around a positively selected variant to decrease. The exact size of the swept region depends on the strength of selection and the speed at which the selected variant rose in frequency. For example, a variant that confers a selective advantage of 1% (considered relatively strong selection) is estimated to sweep a region of approximately 600 kilobases (395). New mutations will eventually re-establish diversity, but these will be slow and infrequent. Positive selection can therefore be observed as a region with low genetic variation that has an excess of rare mutations. Given it takes a neutral variant ~1 million years to fix in the human genome (168), this signature can mark selection that has been acting for hundreds of thousands of years. Unlike the Ka/Ks tests, which involves comparisons between species, identifying reductions in genetic variation is found by comparing diversity within a species. Tests that detect this second signature of selection include the **Tajima's D test** (396) and the **Hudson-Kreitman-Aguadé (HKA) test** (397).

7.3.3 Increase in derived alleles (timescale: < 80 KYA)

When neutral mutations appear in the genome, they will be at a much lower frequency than existing (ancestral) alleles. If, however, the new (or 'derived') mutations are linked to a variant that is positively selected for, these new mutations can also rise in frequency via a selective sweep. The third signature of selection is therefore the presence of many high-frequency derived alleles. It should not be confused with the excess of rare mutations (i.e. low-frequency) mentioned in the previous section. This signature is commonly measured with a **Fay and Wu's H test** (398). Similar to the Ka/Ks test, and unlike the Tajima's D/HKA tests, it involves comparison with the ancestral state – which is typically inferred from a closely related species. Importantly, an increase in derived alleles can only signify selection that has

occurred relatively recently (< 80,000 years ago), as high-frequency derived variants tend to drift to fixation rapidly (399).

7.3.4 Differences between populations (timescale: < 50 – 75 KYA)

This fourth signature also identifies more recent selection, but by comparison *within* a population. For example, human populations in different parts of the world can be subject to markedly diverse selection pressures. An allele that is selected for in one population, but not another, can end up with a measurable difference in frequency between the two gene pools (or with a difference in the variation surrounding the allele). Since such a signature would require populations to be separated for some time (i.e. since the major out-of-Africa migrations), it typically represents selection from 50,000 – 75,000 years ago. Commonly used tests to identify this fourth signature include the **F_{ST} statistic** (400).

7.3.5 Long haplotypes (timescale: <30 KYA)

A **haplotype**, simply put, is a grouping of linked alleles within a chromosomal region (168). Under positive selection, swept regions can rise in frequency so rapidly that recombination has not yet separated the associated alleles. Such undisrupted, or 'long', haplotypes can extend much farther than expected under neutrality. A neutral allele would require a significant amount of time to drift to a high frequency, meaning most common variants are presumed to be old. On the other hand, an allele that has long-range associations with other alleles would be expected to be new (and therefore at a low frequency) since these associations have not been broken by recombination. High frequency variants that are contained in a long haplotype are therefore indicative of recent positive selection. For example, the *LCT* allele that causes lactase persistence is common in Europeans (at ~77% frequency) but lies in a haplotype that extends almost 1 million base pairs, which would not be expected under neutrality (401). This signature is useful for identifying selection that has occurred relatively recently, as recombination will remove long haplotypes after ~30,000 years (168). A commonly used statistic for this signature is the **integrated haplotype score (iHS)** (402).

Signature of positive selection	*Timescale of selection*	*Commonly used tests*
High proportion of non-synonymous substitutions	Millions of years ago	Ka/Ks, McDonald-Kreitman test
Decrease in genetic diversity	< 250,000 years ago	Tajima's D test, Hudson-Kreitman-Aguadé (HKA) test
Excess of derived mutations	< 80,000 years ago	Fay and Wu's H test
Inter-population differences	< 75,000 years ago	F_{ST} statistic
Increased haplotype length	< 30,000 years ago	Integrated haplotype score (iHS)

Table 7.01. Summary table of the five main signatures of positive selection

Each of the signatures is specific for selection that has occurred during a particular timescale. For example, a high ratio of non-synonymous to synonymous substitutions (signature #1) is created by selection that occurred millions of years ago – before the human-chimpanzee split. On the other hand, increased haplotype length is from more recent selection (<30,000 years ago) and is useful for comparisons within the human species. Table adapted from Sabeti et al. (2006).

7.4 Results of the *CNTNAP2* neutrality tests

In light of the evidence of positive selection at the *CNTNAP2* locus (40, 41), I next wanted to perform neutrality tests on a greater number of human populations. As mentioned, the previous neutrality tests only included data from three populations (North-west European, East Asian, and Nigerian). To capture a more diverse set of genomes, I turned to the **Estonian Biocentre Human Genome Diversity Panel (EGDP)** (283). This dataset consists of 369 high-coverage genomes from 12 populations around the globe (see **Table 7.02**). Investigating a larger number of human populations would widen our understanding of the selective forces acting on *CNTNAP2*.

I used two established approaches to test for selection: **1) Tajima's D** (see **Section 7.4.2**) and **2) the integrated haplotype score (iHS)** (see **Section 7.4.5**). As previously described, Tajima's D looks for reductions in genetic diversity and detects selection from $<$ 250,000 years ago (encompassing the origins of modern humans) (50). iHS, conversely, looks for unexpectedly long haplotypes and detects more recent selection ($<$ 30,000 years ago, or after the out-of-Africa migrations). As such, my analysis would capture multiple signatures of selection across both recent and distant human timescales. Under the assumption that *Homo sapiens*-specific traits should be shared by all human populations, any signature of selection operating at the species-level should be observed in distant timescales in all 12 populations. A signature could also be interesting if it is found in African populations (whose genomes were not affected by the Out-of-Africa bottlenecks) (63). If selection is observed in only a few populations with iHS (and not with Tajima's D), then this would suggest selection at the population-level rather than being informative of the evolution of traits uniquely fixed in all humans.

Abbreviation	*Population*	*No. of genomes*
AFR	West and central Africa	26
CSI	Central Siberia	31
ENE	East and north Europe	53
NSI	Northeast Siberia	25
SEA	Island southeast Asia	45
SEM	East and southeast mainland Asia	29
SOA	South Asia	28
SSI	South Siberia and Mongolia	34
SWE	South and west Europe	32
VOL	Volga and Ural	23
WAA	West Asia and Armenia	26
WSI	Western Siberia	17

Table 7.02. Summary chart of the 12 EDGP populations

369 high-coverage genomes from 12 populations were used for the Tajima's D tests. These genomes were taken from the Estonian Biocentre Human Genome Diversity Panel (EGDP) (Pagani et al. 2016). The populations will be referred to by their abbreviations in subsequent figures.

7.4.2 An introduction to the Tajima's D test

The Tajima's D test is a two-tailed test designed to reveal cases of unusually low level of pairwise differences between individual sequences (negative values of the test, indicative of positive selection) and levels of unusually high pairwise differentiation (positive values, characteristic to balancing selection). Calculating Tajima's D involves the following metrics which are illustrated in **Figure 7.01** for clarity (63):

1) θ, the theoretical equilibrium level of genetic variation maintained in a population (without selection or migration, and with a constant population size and random mating). Under neutral conditions, the forces that randomly generate new variants (mutation) and that randomly remove variants (genetic drift) equal each other out to reach what is called the **mutation-drift equilibrium**. Under the equilibrium $\theta = S/Sum(\frac{1}{i})$ where i refers to the number of sequences being sampled.

2) π, the **average** number of **pairwise differences** in the locus of interest that exist in the population. This parameter can be thought of as the observed value of genetic diversity whereas θ is the expectation. Under neutral conditions $\pi = \theta$.

3) S, the number of segregating sites in the sequence of interest (i.e. sites that differ in individuals of the population).

4) n, the number of sequences being compared.

Tajima's D is calculated from:

$$D = \frac{\pi - \theta}{\sqrt{V(\pi - \theta)}}$$

In other words, Tajima's D is simply a normalized measure of the difference between observed and expected genetic diversity ($\pi - \theta$). Normalization is performed by dividing with the standard deviation of that difference (recall SD $= \sqrt{variance}$). If $\boldsymbol{D > 0}$ then there is more genetic diversity than expected (signifying either **balancing selection** at the locus or a **population bottleneck**). If $\boldsymbol{D < 0}$ than there is less diversity than expected, and either **positive selection** or a sudden **population expansion** has occurred. Lastly, if $\boldsymbol{D = 0}$ than the population is at mutation-drift equilibrium (i.e. operating under neutrality) and there is no evidence for selection.

Position	1	2	3	4	5	6	7	8	9	0
Individual A	0	0	0	0	0	0	0	0	0	0
Individual B	0	0	1	0	0	0	0	0	0	0
Individual C	0	0	0	0	0	0	0	0	0	0
Individual D	0	0	0	0	0	0	1	0	0	0
Individual E	0	0	0	0	0	0	1	0	0	0

Pairwise differences

A vs. B: 1 difference
A vs. D: 1 difference
A vs. E: 1 difference
B vs. C: 1 difference
B vs. D: 1 difference
B vs. E: 1 difference
C vs. D: 1 difference
C vs. E: 1 difference

Calculation of π:

$$\pi = \frac{\sum_{i<j} d_{ij}}{n(n-1)/2}$$

$$\pi = \frac{1+1+1+1+1+1+1+1}{5(5-1)/2}$$

$$\pi = 0.8$$

Calculation of θ:

$$\theta = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

$$\theta = \frac{2}{\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}}$$

$$\theta = 0.96$$

Figure 7.01. Examples of π and θ

An example locus is shown for 5 individuals across 10 nucleotide positions. Shared nucleotides are denoted by '0' and differences by '1'. There are 8 pairwise differences, d_{ij}, occurring at 2 different positions, *S*. These parameters are used to calculate the observed variation, π and compare it to the expected variation, θ. Under neutrality $\pi = \theta$. If positive selection has occurred then $\pi > \theta$, and if balancing selection has occurred then $\pi < \theta$.

7.4.4 Results of the CNTNAP2 Tajima's D test

To evaluate signals of positive and balancing selection at the *CNTNAP2* gene, Tajima's D was calculated across chromosome 7 (which contains *CNTNAP2*) in 2 kb non-overlapping windows (~160 Mb, ~80,000 windows)[9]. The 2 kb window size was chosen to accurately detect long-term balancing selection, which is thought to leave signatures of only a few kilobases (403, 404). As mentioned, reference genomes were taken from the Estonian Biocentre Human Genome Diversity Panel (EGDP) (**Table 7.02**) (283).

Table 7.03 reports the average Tajima's D statistic for the windows containing the *CNTNAP2* HARs. A window was considered potentially significant if it fell within the top 1% most positive or top 1% most negative chromosome-wide for any of the 12 populations (see **Figure 7.02**). Most of the *CNTNAP2* HARs had Tajima's D values close to zero, indicating no selection across the populations. HACNS_116 and HACNS_97, however, tended to show negative values. Specifically, HACNS_116 had three populations within the top 5% and HACNS_97 had two in the top 1% [**HACNS_116 –** CSI, SEA, and WSI; **HACNS_97 –** SEM and SSI]. 2xHAR.395 was in the top 5% chromosome-wide in one population, NSI. As such, there is no strong evidence of human-wide positive selection at any of the *CNTNAP2* HARs. There is also no evidence of positive selection at the HARs in African populations (AFR). These data do, however, suggest some of the HACNS_116 and HACNS_97 may be positively selected for in certain populations, but these signals are not particularly strong (e.g. they may fall out of significance genome-wide).

Looking outside of the HARs, and across the *CNTNAP2* gene more generally, there is a clear enrichment of negative D values in introns 1 and 13 (as previously reported (40, 41)). Given that intron 13 overlaps a hot-spot for mutations associated with specific language impairment (SLI) (see **Figure 2.08**), and that intron 1 contains

[9] With thanks to Sarah Kaewert for providing her code and technical support.

half of the gene's HARs, these findings are still noteworthy. They will be contextualized further in the discussion section of this chapter.

Perhaps more interesting, was the finding of extremely **positive** values at ANC1208 and ANC1209 - in a number of the populations. ANC1208 ranked in the top 1% most positive windows in three populations [NSI: top 638 bins, SEA: top 754 bins, and WSI: top 271 bins], and within the top 5% in four populations (CSI, ENE, SEM, and WAA). Even more striking were the results for ANC1209, which ranked in the top 1% most positive in eight of the twelve populations [CSI: top 41 bins, ENE: top 85 bins, SOA: top 71 bins, SSI: top 30 bins, SWE: top 53 bins, VOL: top 121 bins, WAA: **top 5 bins**, and WSI: top 588 bins]. Of the remaining four populations, it was in the top 5% in three (**AFR**, SEA, and SEM). Since the HAR method supposedly identifies human-specific positive selection, it was highly unexpected to find evidence of balancing selection at these loci. It was perhaps even more surprising to find evidence of this in the African genomes. This supposed paradox will also be discussed in this chapter's discussion section.

Once more, zooming out and analyzing the entire *CNTNAP2* locus reveals several interesting patterns. Firstly, there appears to be a concentration of positive D values in introns 1, 11, 13, and 20. Secondly, the top 1% most positive African values are visibly lower than for the other 11 populations. This is likely explained by the fact that the EGDP African genomes were a pool of several different sub-populations (283). In that way, they may be cancelling each other out to reduce the detected signal strength. (All other populations in the EGDP dataset are distinct, and selection signatures should not be diluted out).

HAR	*AFR*	*CSI*	*ENE*	*NSI*	*SEA*	*SEM*	*SOA*	*SSI*	*SWE*	*VOL*	*WAA*	*WSI*
HACNS_116	-0.463	-1.705	-0.480	-1.103	-1.656	-1.138	-1.012	-0.488	-0.044	-1.116	0.122	-1.622
HACNS_97	-0.315	nan	-1.446	nan	-1.040	-1.976	-1.369	-1.959	-1.320	-1.448	-1.609	nan
2xHAR.395	0.187	0.912	1.028	-1.782	0.150	0.552	0.388	0.538	1.165	1.155	-0.093	0.244
HACNS_884	-1.266	0.749	-1.003	-0.450	-1.042	-0.088	-0.895	1.246	-0.364	-0.015	-0.622	-0.075
ANC1208	0.135	2.247	2.250	2.794	2.706	1.894	-0.150	1.701	0.784	1.393	2.014	2.784
HACNS_590	0.210	-0.029	-0.425	0.307	0.258	0.817	-1.267	-0.400	-1.124	-0.211	0.429	-0.744
ANC1209	0.829	3.546	3.314	0.978	2.193	2.384	3.120	3.466	3.301	2.997	3.632	2.560
HACNS_954	0.618	-0.426	0.847	2.507	-1.080	0.355	0.716	0.610	0.348	1.039	0.302	0.692
CNTNAP2 avg	-0.640	0.001	0.041	0.329	-0.081	-0.039	-0.290	-0.115	0.055	-0.012	-0.125	0.057

Table 7.03. Tajima's D for 2 kb windows encompassing the *CNTNAP2* HARs

The Tajima's D values reported are the average value for each population (see **Table 7.02** for abbreviations and sample sizes). While most HARs had D values around zero (indicating no selection), HACNS_116 and HACNS_97 showed modestly negative measures while ANC1208 and ANC1209 showed positive values. In 11 of the 12 populations examined, ANC1209 ranked in the top 5% most positive windows across chromosome 7. ANC1208 ranked in the top 5% most positive in seven of the populations. Blue shading denotes a Tajima's D value is within the top 5% most negative windows of chromosome 7. Red shading indicates it is in the top 5% most positive.

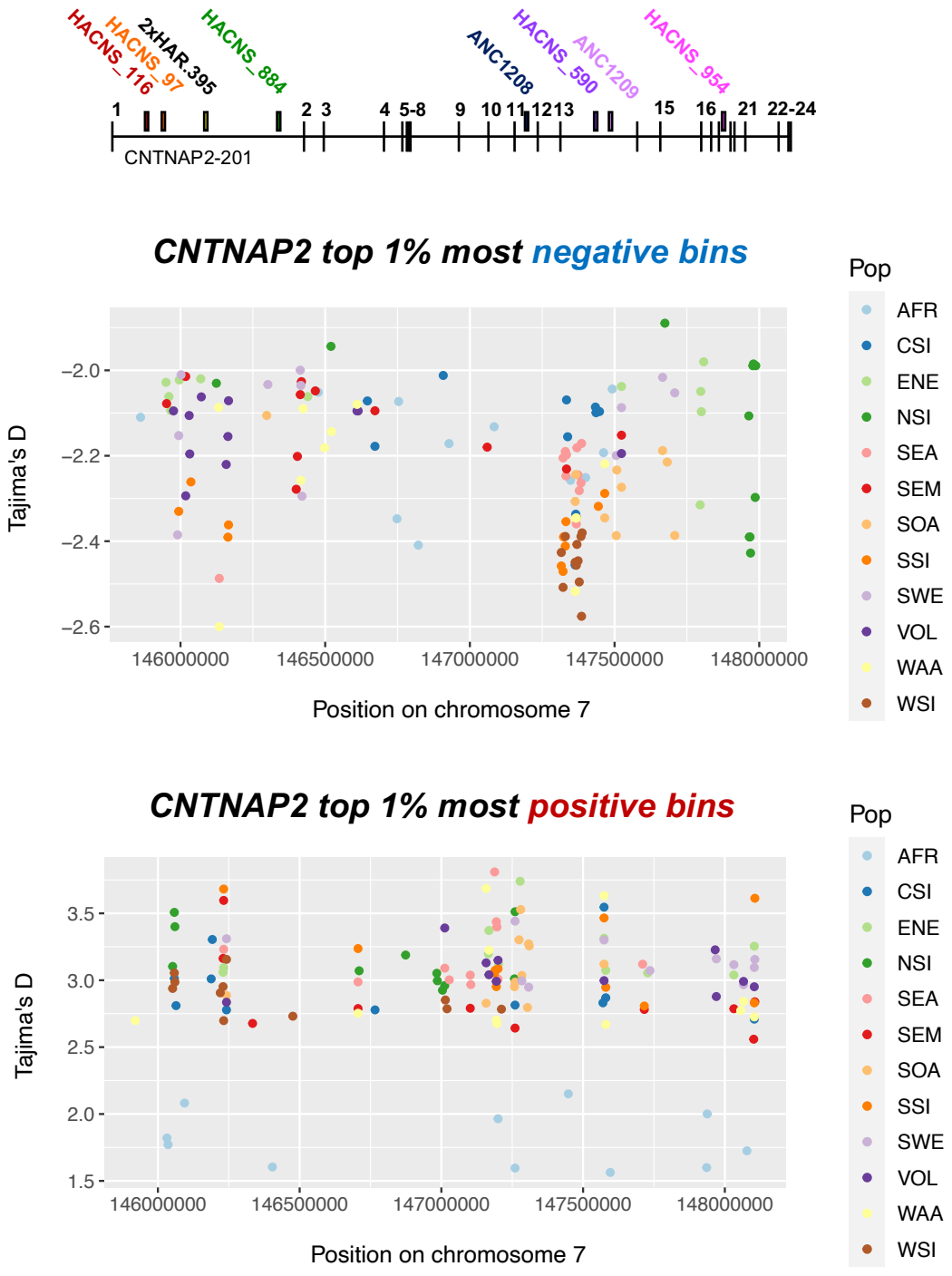


Figure 7.02. The top 1% Tajima's D of the *CNTNAP2* gene

For each of the 12 populations, the top 1% most positive and most negative Tajima's D statistics of the *CNTNAP2* gene are plotted. Strongly positive values could indicate either balancing selection or a population expansion, while strongly negative values could indicate positive selection or a population bottleneck. A schematic of the *CNTNAP2* locus is shown at the top of the page, to be used as a scaled reference for the x-axis of each graph.

7.4.5 An introduction to the integrated haplotype score (iHS)

As explained in **Section 7.3.5**, high frequency variants that are contained in a long haplotype are indicative of recent positive selection (168). The iHS test (402) is a frequently used approach to test for this signature. The test is based on another metric called the **extended haplotype homozygosity (EHH)** statistic (405). EHH measures the decay of haplotype homozygosity from a **'core SNP'**. The algorithm assigns the observed homozygosity a score from 1 to 0 with decreasing values as distance from the core SNP increases. Under neutrality, newly arising derived alleles (i.e. created by mutation) will either be removed by drift or rise slowly in frequency. Recombination over time will break the linkage between physically separated loci with the result that high frequency alleles will be associated with short haplotypes. Therefore, under neutral scenarios it is uncommon to find alleles with high frequency and long surrounding haplotype homozygosity. When an allele rapidly rises in frequency due to positive selection, its haplotype homozygosity is expected to extend much further than under neutrality (402). Consequently, in plots of EHH versus distance from the core, the area under the EHH curve will be greater for a selected allele than for a neutral allele.

Voight et al. (402) translated these EHH statistics into their iHS scores. Specifically, they computed the integral of the decay of EHH away from a specified core allele until EHH reached 0.05. This **integrated EHH (iHH)** was summed over both directions away from the core SNP. They then obtained their iHS statistic by taking the natural log of the iHH for the ancestral core allele (iHH_A) versus the iHH of the derived core allele (iHH_D):

$$unstandardized\ iHS = ln\left(\frac{iHH_A}{iHH_D}\right)$$

If the rate of EHH decay is similar on the ancestral and derived alleles then $iHH_A/iHH_D = 1$ and the unstandardized iHS = 0. Large negative values indicate the derived allele is in an unusually long haplotype for its frequency. Conversely, large positive values indicate the ancestral allele is in a long haplotype. The unstandardized iHS is adjusted to obtain the final **standardized iHS** statistic which has a mean = 0 and a variance = 1 regardless of the frequency of the core SNP:

$$standardized\ iHS = \frac{ln\left(\frac{iHH_A}{iHH_D}\right) - E_p\left[ln\left(\frac{iHH_A}{iHH_D}\right)\right]}{SD_p\left[ln\left(\frac{iHH_A}{iHH_D}\right)\right]}$$

The **expectation (E)** and **standard deviation (SD)** of $\ln(iHH_A/iHH_D)$ are estimated from the empirical distribution of SNPs with the same **derived allele frequency (p)** as the core SNP. Since the iHS is constructed to be normally distributed, the sizes of iHS signals from different SNPs are directly comparable (regardless of their allele frequencies). Moreover, since iHS is standardized using genome-wide empirical distributions, it also gives an indication of how unusual the haplotypes around a given SNP are. Since selective sweeps should cause clusters of extreme iHS scores across the swept region, it is more powerful to look for windows with numerous extreme iHS statistics than to examine SNPs independently. Under neutrality, extreme iHS scores would be scattered uniformly across the genome (unpublished modeling from Voight et al. (402)).

Large negative iHS scores, in principle, should indicate positive selection. However, in simulations Voight et al. (402) noted selective sweep regions can also produce large positive iHS values if ancestral alleles have been swept along with the selected site. Furthermore, it is possible that selection may switch to favor an ancestral allele that has been segregating in the population. For these reasons, both extreme positive and extreme negative iHS scores should be considered as potential signs of positive selection.

7.4.6 Results of the integrated haplotype score (iHS)

iHS scores were previously computed for the EGDP dataset by members of the Kivisild group (data and detailed methods are published in Pagani et al. (283)). Scores were analysed for 200 kb non-overlapping windows. Each window was ranked according to the proportion of SNPs with |iHS|>2. These ranks were then translated into empirical *p*-values by dividing by the total number of windows in the genome (~13,000).

12 windows overlapped the *CNTNAP2* locus with six of them containing HARs (**Table 7.04**). While no window was considered significant genome-wide ($p > 0.05$ for all), there were two windows that ranked in the top 10% genome-wide. **Chr7_145.8 Mb** ranked in the top 6.8% in Africa, top 8.8% in South/West Europe, and top 7.4% in Volga/Ural. **Chr7_146.4 Mb** reached the top 8.9% in South Asia and top 5.8% in West Asia/Armenia. The first of the two windows, Chr7_145.8 Mb, contains HACNS_116 and HACNS_97. The second window did not overlap any of the *CNTNAP2* HARs.

Window	*HARs*	*AFR*	*CSI*	*ENE*	*NSI*	*SEA*	*SEM*	*SOA*	*SSI*	*SWE*	*VOL*	*WAA*	*WSI*
7_145.8	2	0.068	0.365	0.719	0.556	0.676	0.841	0.322	0.849	0.088	0.074	0.464	0.634
7_146	1	0.549	0.848	0.714	0.360	0.893	0.908	0.717	0.915	0.755	0.549	0.225	0.102
7_146.2	1	0.171	0.160	0.903	0.800	0.893	0.825	0.215	0.898	0.382	0.911	0.442	0.855
7_146.4	-	0.322	0.724	0.304	0.457	0.839	0.830	0.089	0.582	0.400	0.404	0.058	0.440
7_146.6	-	0.315	0.648	0.784	0.703	0.547	0.613	0.607	0.202	0.889	0.422	0.566	0.657
7_146.8	-	0.873	0.848	0.393	0.192	0.651	0.833	0.476	0.198	0.574	0.418	0.475	0.814
7_147	-	0.893	0.752	0.823	0.176	0.893	0.494	0.758	0.914	0.541	0.671	0.932	0.579
7_147.2	1	0.590	0.155	0.918	0.184	0.414	0.474	0.556	0.388	0.603	0.490	0.932	0.133
7_147.4	2	0.607	0.747	0.272	0.809	0.278	0.161	0.337	0.115	0.565	0.460	0.858	0.344
7_147.6	-	0.520	0.843	0.669	0.508	0.821	0.757	0.366	0.809	0.244	0.561	0.633	0.489
7_147.8	1	0.540	0.705	0.659	0.737	0.573	0.849	0.540	0.915	0.246	0.734	0.639	0.870
7_148	-	0.622	0.142	0.318	0.788	0.682	0.372	0.850	0.206	0.321	0.865	0.523	0.166

Table 7.04. iHS significance scores for 200 kb windows encompassing the *CNTNAP2* locus

For each 200 kb window in the *CNTNAP2* locus (and each of the 12 EGDP populations) an empirical *p*-value is reported. *P*-values were calculated by ranking each of the 200 kb windows in the human genome (~13,000) by the number of SNPs they contain with a standardized iHS score > 2 or < -2. Windows with *p*-values in the top 10% genome-wide are shown in red. Also indicated are the number of human accelerated regions (HARs) contained within each 200 kb bin. Coordinates map to GRCh37 and are shown in the form of Chromosome_Mb_start (e.g. for the first row, the window begins at position 145.8Mb on chromosome 7).

7.5 Purifying selection at the *CNTNAP2* HARs

7.5.1 The Combined Annotation-Dependent Depletion (CADD) Score

Negative selection implies a locus is under constraint and is therefore potentially functional. As such, it is another important mode of selection to test the *CNTNAP2* locus for. One easily implemented approach is to use the **Combined Annotation-Dependent Depletion (CADD)** algorithm (284, 285). CADD computes a 'deleteriousness score' for any of the three possible base changes relative to the human reference genome (a proxy for negative selection). The score is generated by a machine learning algorithm that combines 63 annotations into a single metric. These annotations include **1) disease mutation databases**, **2) variant effect predictions**, **3) conservation estimates (e.g. phastCons)**, **4) epigenetic marks**, **5) expression in common cell lines**, and **6) surrounding sequence context** (e.g. distance to exon-intron boundaries). To improve interpretability, raw CADD scores are transformed into a 'PHRED-like' score (i.e. log10) based on each variant's deleteriousness relative to all possible SNVs (for which there are ~9 billion in the human genome). A PHRED-like CADD score of 10 or more indicates the raw score was in the top 10% most deleterious of all possible SNVs. A score of >20 indicates the raw score was in the top 1% most deleterious, >30 in the top 0.1% most deleterious, and so on.

CADD scores were first introduced in 2014 to improve existing variant prediction methods (e.g. PolyPhen, SIFT, and others) (284). Most previous approaches were based on conservation metrics that did not incorporate functional data. Others used missense scoring tools which cannot be used on non-coding DNA, thereby excluding >99% of human genetic variation. Crucially, CADD is also more effective at correctly predicting deleteriousness. For instance, it accurately ranks known disease-causing mutations (ClinVar pathogenic variants) as harmful (284). PolyPhen, on the other hand, misses approximately 27% of them. CADD assignments have also been shown to correlate with GWAS results, offering further support for

their usefulness (284). In light of these benefits, I next used CADD to predict the harmfulness of potential mutations at each of the eight *CNTNAP2* HARs.

7.5.2 CADD scores of the CNTNAP2 HARs

CADD scores were computed for all possible SNVs at each of the *CNTNAP2* HARs (plus 1 kb upstream and downstream for context) (**Figure 7.03**). Six of the HARs overlapped with local peaks in deleteriousness (HACNS_116, HACNS_97, HACNS_884, ANC1208, HACNS_590, and HACNS_954). The sequences contained CADD scores near or above 20, indicating mutations in these regions would be within the top 1% most harmful genome-wide. Outside of the HAR boundaries, CADD scores tended to drop from a maximum of ~20 to ~10 (apart from HACNS_954, where an exon-intron boundary caused scores to spike over 40). This suggests a fraction of the *CNTNAP2* HARs could perhaps be discrete, functional elements within a larger intronic environment. It is also worth mentioning that at a given nucleotide position, CADD scores did not vary significantly between putative changes. A 'C'-nucleotide changed to an 'A', 'G', or 'T' generally all had similar scores. Therefore, the act of mutation itself appeared to be more important than the precise nucleotide change.

Table 7.05 reports the mean, maximum, and minimum CADD scores for each HAR, along with the percentage of variants with scores ≥ 20. Also reported are the same metrics for chromosome 7. Of the eight *CNTNAP2* HARs, HACNS_97 had the highest individual CADD score (22.1), while HACNS_590 had the highest mean score (15.5). Both HACNS_97 and HACNS_590 appeared enriched for variants with CADD scores in the top 1% genome-wide – 12.2% and 12.7% of all possible mutations respectively. Of the remaining possible substitutions, 58% in HACNS_97 and 88% in HACNS_590 contained scores in the top 10% most deleterious (i.e. scores ≥ 10). These results were notably higher than the mean and minimum scores for chromosome 7 (4.31 and 0.001, respectively). However, the chromosome-wide maximum score (84) was considerably larger than any of the maximum HAR values. This is almost certainly due to the fact that many coding regions, promoters, and known functional elements have higher CADD scores than non-coding DNA (discussed further below) (284).

In addition to the HARs that were enriched for high CADD scores, three others – 2xHAR.395, HACNS_884, and ANC1209 – had less noteworthy metrics. None contained a single variant with a score over 20, and all three had generally quite low maximums (e.g. 13.7 for 2xHAR.395). It was interesting to find ANC1209 had relatively unremarkable CADD scores given the signs of balancing selection at this region. Under balancing selection, one could rationalize finding more mutations benign than under positive selection. Having said that, since balancing selection is still a form of adaptive constraint, I would not expect all substitutions to have low CADD scores.

Also shown in **Figure 7.03/Table 7.05** are the CADD scores for a human forebrain enhancer in the *GLI3* gene (validated in the Vista Enhancer Browser (406) and independently in Paparidis et al. (407)). As evident, the scores for this enhancer are strikingly high with 42% of all variants in the region scoring over 20 (~808/1911 variants). The maximum score – at 24.3 - was also higher than detected in any of the *CNTNAP2* HARs. From this point of view the *CNTNAP2* HARs may not be as functional or as significant as the *GLI3* enhancer. There are also other possible explanations for this disparity – these will be analyzed further in the **Discussion** section.

To put these findings into perspective, it is important to stress a few points. Firstly, the vast majority of variants with high CADD scores fall in protein-coding portions of the genome (284). Of the ~90 million changes possessing a CADD score over 20, only 22% are intronic - or approximately 21 million of them. The highest CADD score calculated for an intronic mutation is 39, for which there is a single variant identified (284). It is exciting, therefore, to find that some of the HARs contain high scoring variants. That said, it needs to be clearly emphasized that these findings are not unusual in the absolute sense – there are millions of other variants with similar estimated levels of deleteriousness. What CADD scores do provide, is additional evidence that purifying selection may be acting, and that the *CNTNAP2* HARs should not be immediately dismissed as non-functional.

Having observed several of the HARs overlapped high scoring variants, I wondered whether these were enriched for mutations at the sites of human-specific nucleotide changes. In particular, I was curious about whether mutations back to the nucleotide state in chimpanzees would amass particularly high scores. In **Figure 7.03**, variants shown in red represent the nucleotide present in chimpanzees. At least for the *CNTNAP2* HARs, it did not appear these sites had particularly noteworthy CADD scores. None scored in the top 1% most deleterious genome-wide, and the highest scores were generally around 15 (top 5% most deleterious). If the HARs were positively selected for, or any of the human-specific nucleotides within them, I would expect a loss of that region/nucleotide to be damaging.

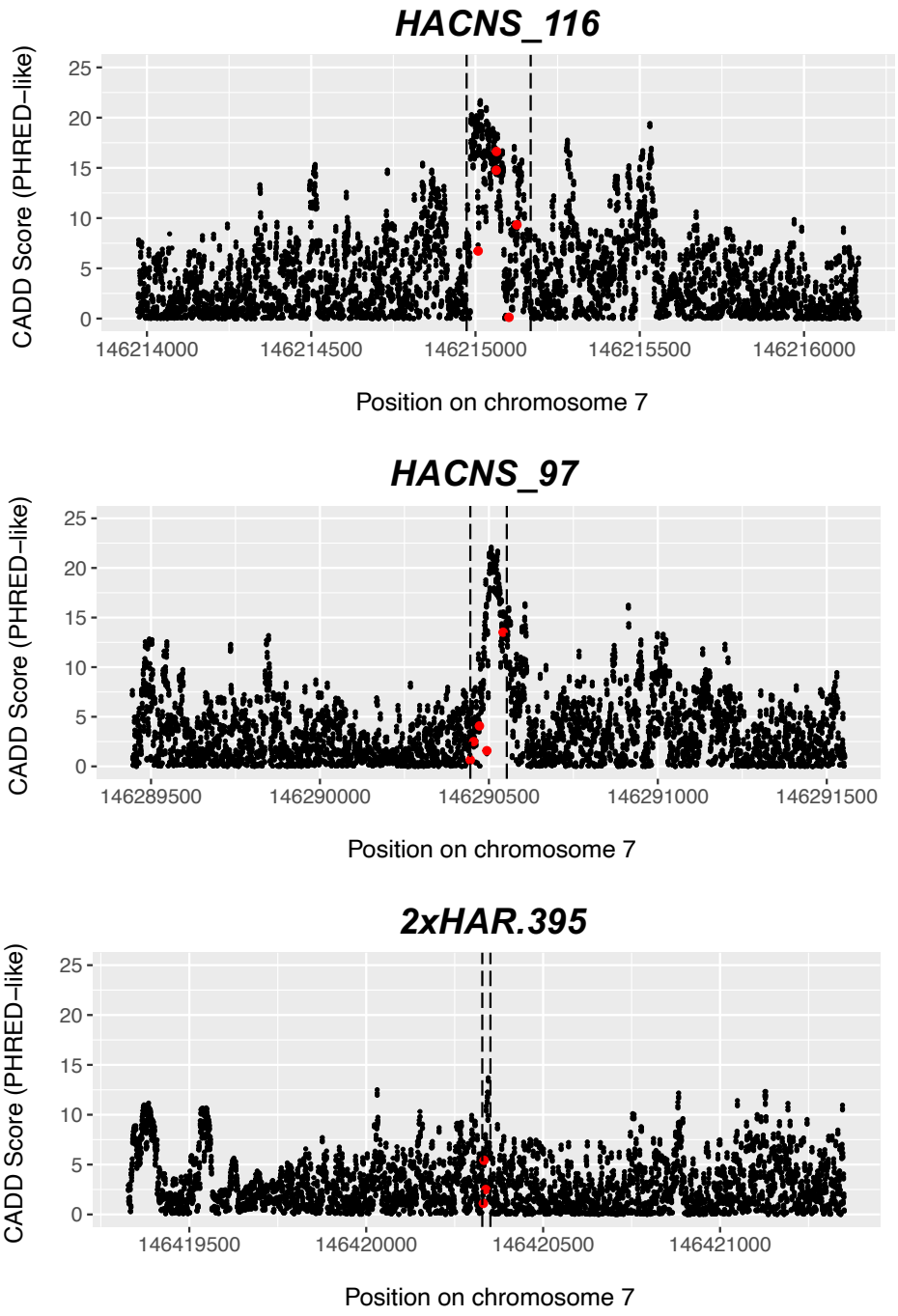


Figure 7.03. CADD scores for each of the *CNTNAP2* HARs (2 kb windows)

Combined Annotation-Dependent Depletion (CADD) scores are plotted for each *CNTNAP2* HAR (plus 1 kb pads on either side). For every nucleotide in the ~2 kb window, all three possible mutations relative to the reference genome are reported. A PHRED-like (i.e. log10) CADD score of >20 indicates the variant is in the top 1% most deleterious in the genome. Dashed lines represent the boundaries of each HAR. Mutations that would revert a human-specific nucleotide to the chimpanzee state are highlighted in red. Most of the HARs overlap local peaks in deleteriousness – except for 2xHAR.395 and ANC1209. *Plots continue on the next page.*

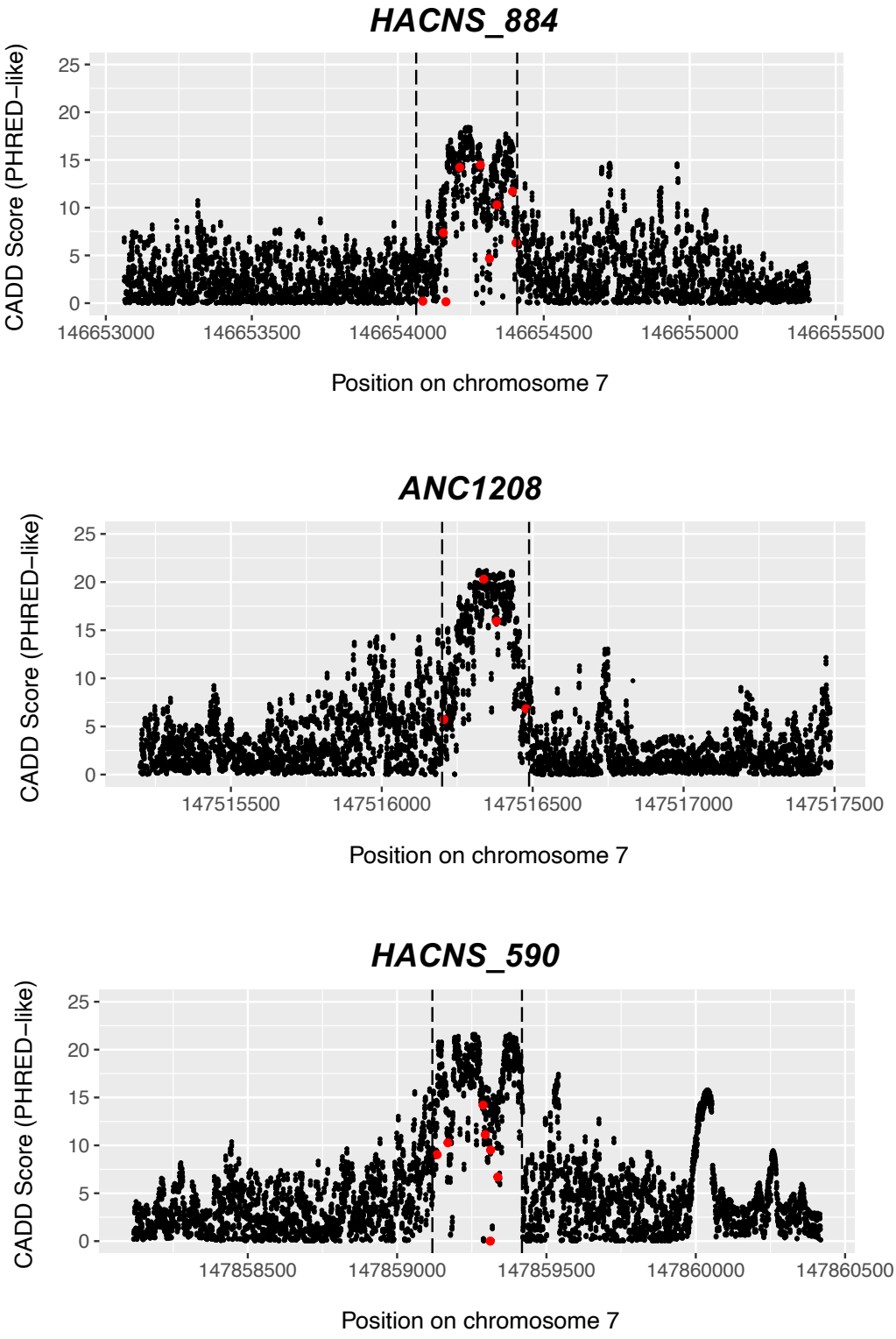


Figure 7.03. CADD scores for each of the *CNTNAP2* HARs (2 kb windows)

Continued from the previous page.

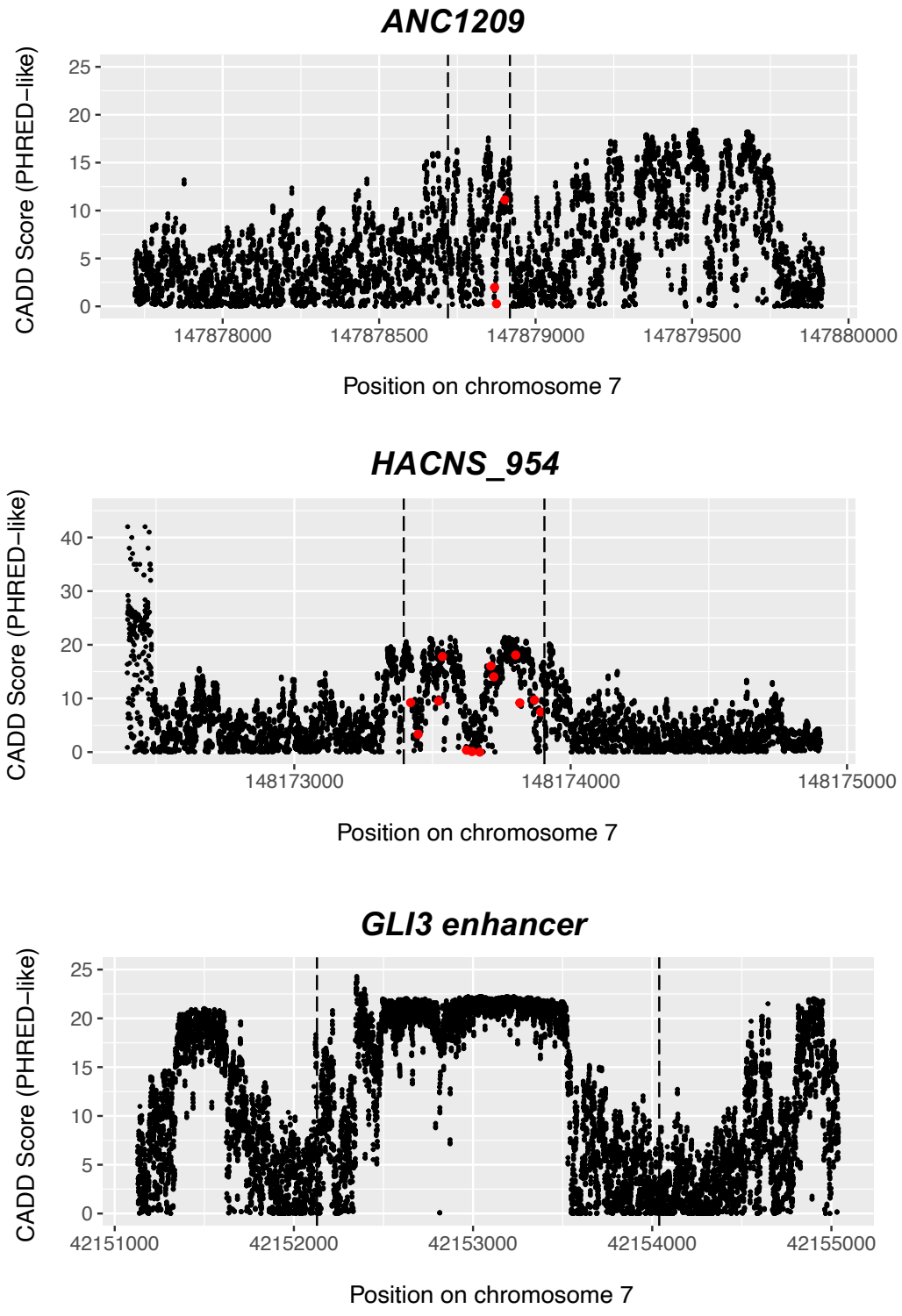


Figure 7.03. CADD scores for each of the *CNTNAP2* HARs (2 kb windows)

Continued from the previous page. Note the y-axis for CNS954 extends to 50 (scores >40 correspond to the exon 18-intron 18 boundary). CADD scores for a validated forebrain enhancer (located in the *GLI3* gene) is also shown for reference.

HAR	*HAR mean*	*HAR max*	*HAR min*	*% of HAR variants ≥ 20*	*Sequence length (bp)*
HACNS_116	11.5	21.7	0.006	3.4%	196
HACNS_97	12.2	22.1	0.03	12.2%	109
2xHAR.395	5.7	13.7	1.1	0%	23
HACNS_884	10.5	18.4	0.02	0%	347
ANC1208	14.1	21.2	0.02	9.2%	289
HACNS_590	15.5	21.6	0.001	12.7%	301
ANC1209	7.7	17.6	0.02	0%	199
HACNS_954	12.5	21.4	0.006	4.6%	510
GLI3 enhancer	14.5	24.3	0.001	42.3%	1911
Chromosome 7	4.31	84.0	0.001	0.88%	159,345,973

Table 7.05. CADD score summary statistics for the *CNTNAP2* HARS

The mean, maximum, and minimum CADD scores from all possible substitutions in each of the HARs are reported. The proportion of substitutions that would fall in the top 1% most deleterious in the genome (CADD score ≥20) is also provided along with the sequence length. HACNS_97 and HACNS_590 appeared somewhat enriched in variants with CADD scores over 20. Statistics for a validated forebrain enhancer (located within intron 2 of the *GLI3* gene) and for all of chromosome 7 are also shown for reference.

7.6 Bioinformatic analyses of *CNTNAP2* HAR enhancer function in the human cortex

7.6.1 ChromHMM chromatin state annotations

In addition to looking for signs of selection at the *CNTNAP2* HARs, I also wanted to clarify whether any of the HARs overlapped epigenetic marks of gene regulatory elements. To investigate this, I used **ChromHMM**, an algorithm that translates multiple epigenetic marks into a chromatin state prediction (408). 15 possible states are considered, ranging from 'quiescent' to 'active transcriptional start site', and are assigned at a 200 bp resolution. For a full description of the 15 states please see **Table 7.06a.** For a summary of the chromatin marks used by ChromHMM please see **Table 7.06b**.

The **NIH Roadmap Epigenomics Mapping Consortium** (409) has published ChromHMM annotations for 127 epigenomes (410). To analyze the enhancer potential of the *CNTNAP2* HARs, I queried the ChromHMM state of each HAR across 14 types of neural tissue from adult, fetal, *in-vivo* and *in vitro* sources (samples are reported in **Table 7.07**). To test the specificity of an enhancer assignment to the brain (should one occur), I also obtained the chromatin states for each HAR in non-neural tissue. ChromHMM annotation tracks were visualized on the **WashU Epigenome Browser** (411).

ChromHMM results are shown in **Figure 7.04a-h**. Interestingly, four of the eight HARs were classified as an 'enhancer' in primary brain tissue: 1) **HACNS_884** in substantia nigra, 2) **HACNS_590** in inferior temporal lobe, 3) **ANC1209** in substantia nigra, and 4) **HACNS_954** in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak

and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites.

A

Chromatin State	*Abbreviation*	*Annotation Colour*
Active transcriptional start site	*TSSa*	
Flanking Active TSS	*TssAFlnk*	
Transcr. at gene 5' and 3'	*TxFlnk*	
Strong transcription	*Tx*	
Weak transcription	*TxWk*	
Genic enhancer	*EnhG*	
Enhancer	*Enh*	
ZNF genes & repeats	*ZNF/Rpt*	
Heterochromatin	*Het*	
Bivalent/Poised TSS	*TSSBiv*	
Flanking Bivalent TSS/Enh	*BivFlnk*	
Bivalent Enhancer	*EnhBiv*	
Repressed PolyComb	*ReprPC*	
Weak Repressed PolyComb	*ReprPCWk*	
Quiescent/low	*Quies*	

B

Mark	*Meaning*	*Effect*
H3K4me3	Tri-methylation of lysine-4 on histone H3	Promoters
H3K4me1	Methylation of lysine-4 on histone H3	Enhancers
H3K36me3	Tri-methylation of lysine-36 on histone H3	Transcribed DNA
H3K27me3	Tri-methylation of lysine-27 on histone H3	Repressive
H3K9me3	Tri-methylation of lysine-9 on histone H3	Repressive

Table 7.06. Summary of ChromHMM chromatin state annotations

(A) ChromHMM integrates multiple chromatin datasets (ChIP-Seq of histone modifications) into a prediction of 15 possible chromatin state annotations. Each state is represented by a uniquely colored annotation track on the WashU epigenome browser. (B) The five epigenetic marks assessed by the ChromHMM algorithm and their effects on transcription.

Sample	*Reference*
H1 Derived Neuronal Progenitor Cultured Cells	*E007*
H9 Derived Neuron Cultured Cells	*E010*
Cortex derived primary cultured neurospheres	*E053*
Ganglion Eminence derived primary cultured neurospheres	*E054*
Brain Hippocampus Middle	*E071*
Brain Substantia Nigra	*E074*
Brain Cingulate Gyrus	*E069*
Brain Angular Gyrus	*E067*
Brain Anterior Caudate	*E068*
Brain Inferior Temporal Lobe	*E072*
Brain Dorsolateral Prefrontal Cortex	*E073*
Brain Germinal Matrix	*E070*
Fetal Brain Male	*E081*
Fetal Brain Female	*E082*
Aorta	E065
Fetal Stomach	E092
Fetal Lung	E088
Fetal Kidney	E086
Fetal Heart	E083
Pancreas	E098
Lung	E096
Esophagus	E079
Liver	E066
Spleen	E113
Placenta	E091
Small Intestine	E109

Table 7.07. Summary of Roadmap epigenomes

Information on the Roadmap samples used to investigate the overlap of enhancer marks with the *CNTNAP2* HARs. Samples from both sexes, across a range of brain regions, and from a number of ages (both fetal and adult) were included. For further sample details please refer to the Gene Expression Omnibus (GEO). GEO accession numbers are listed for each sample.

A

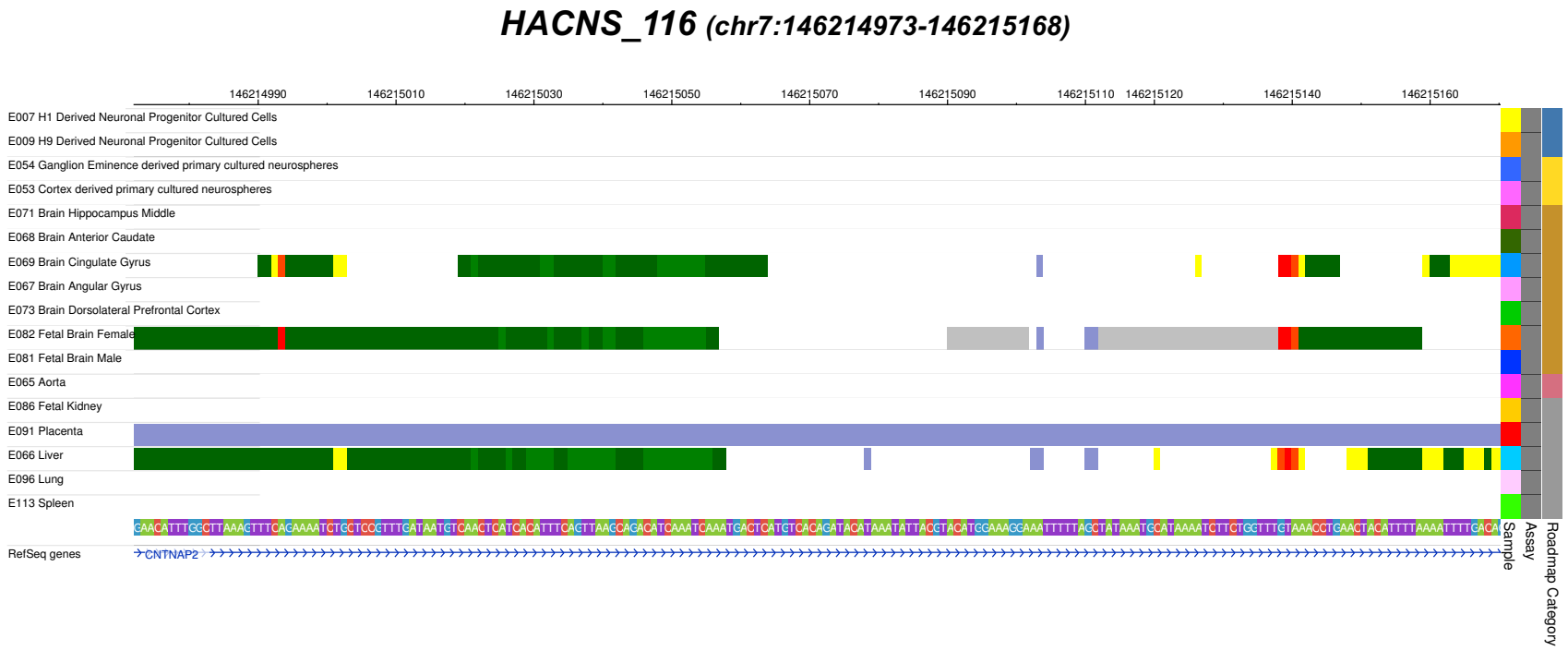


Figure 7.04. ChromHMM annotations for the eight CNTNAP2 HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

B

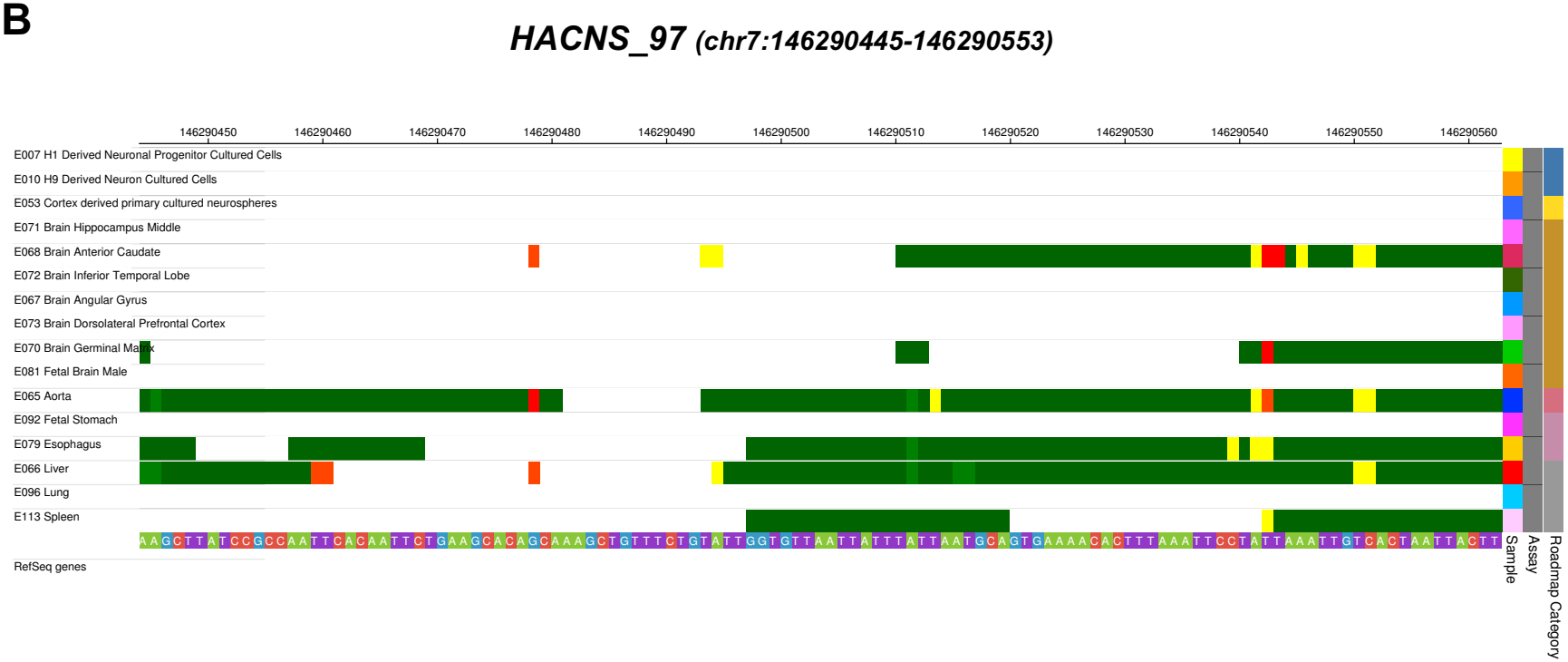


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

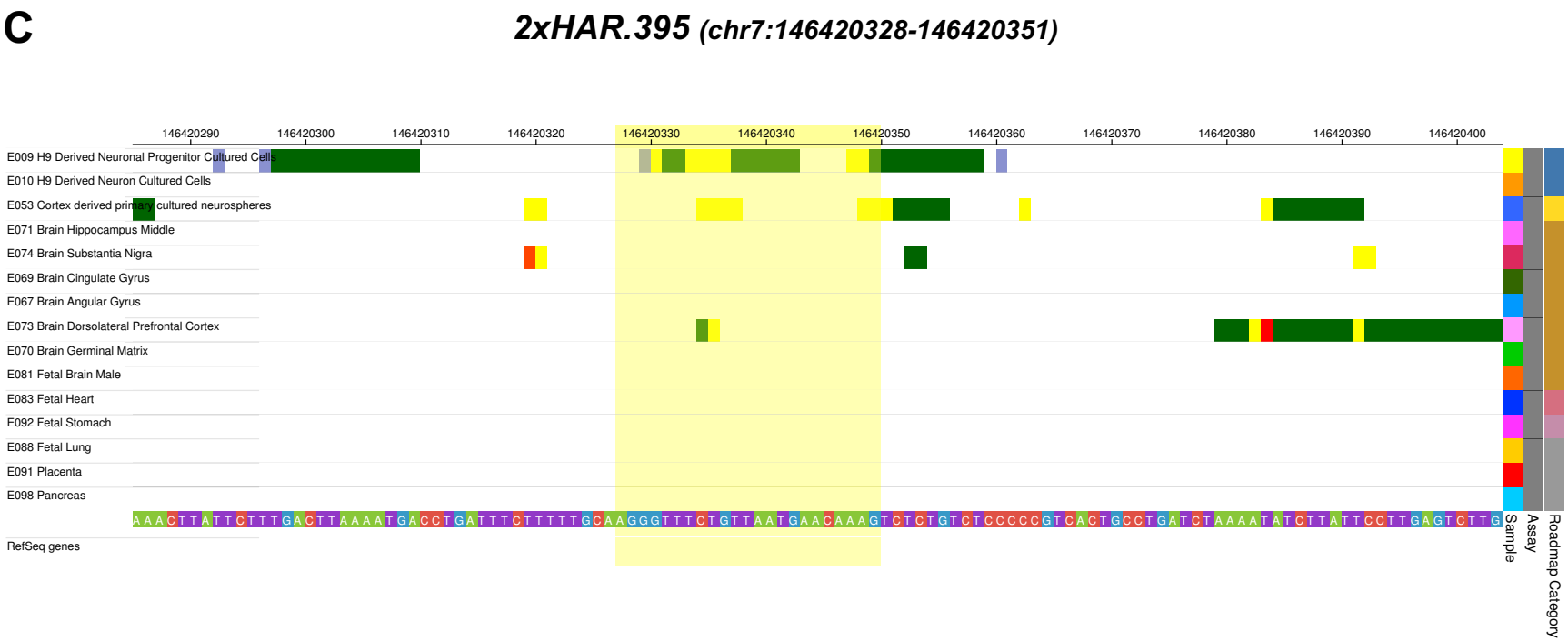


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

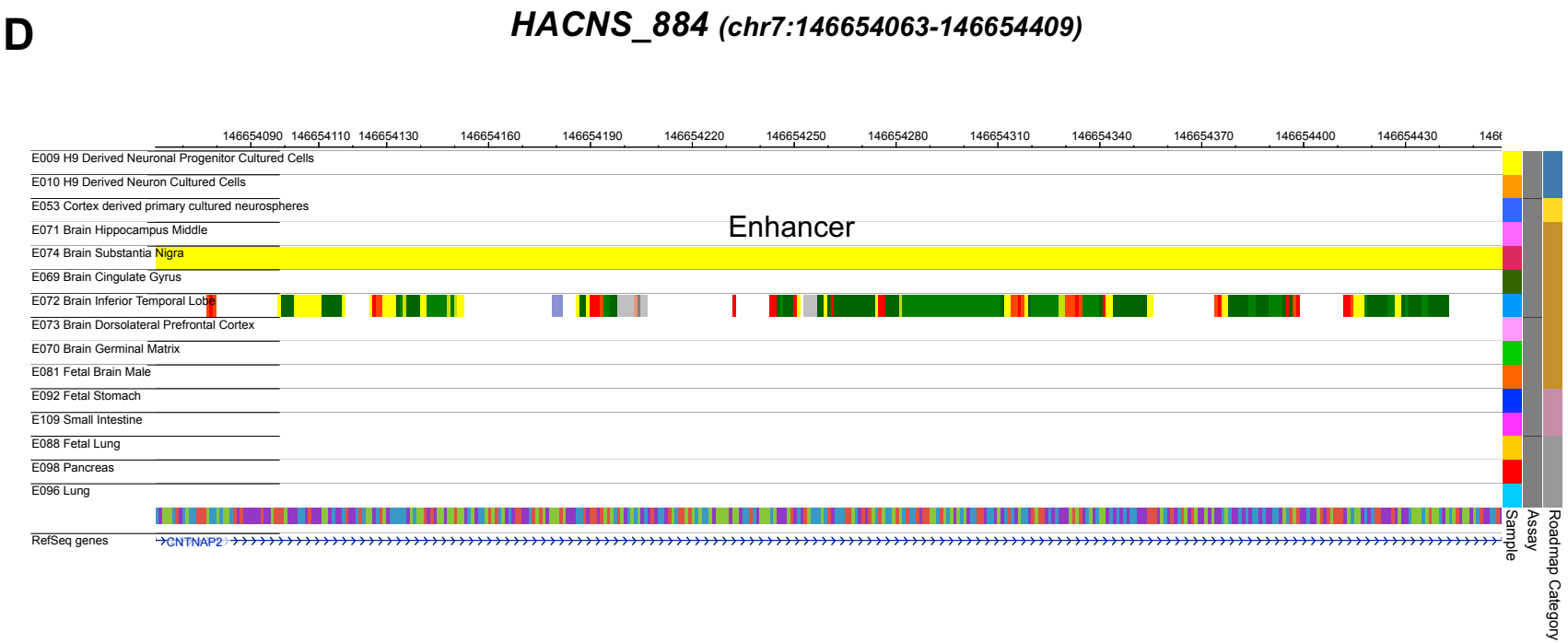


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

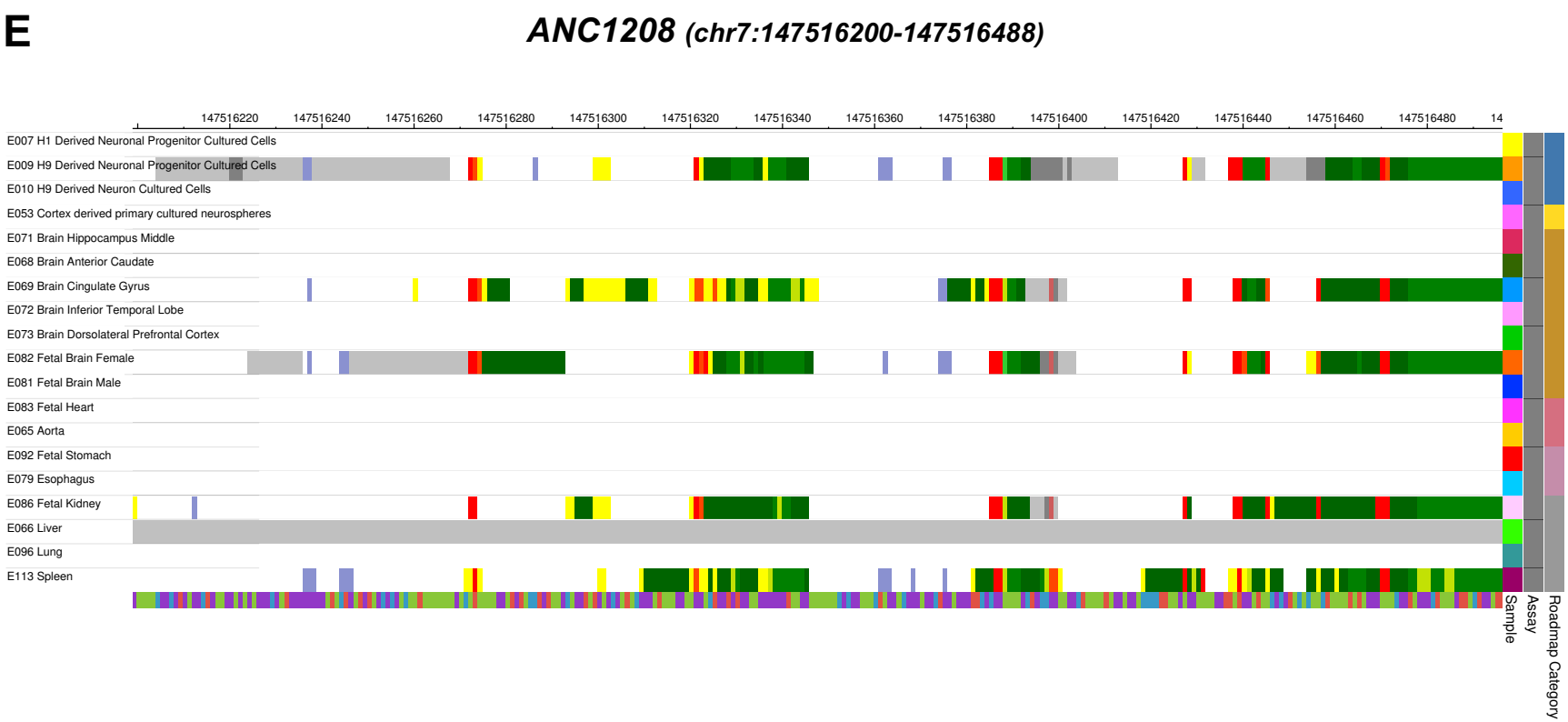


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

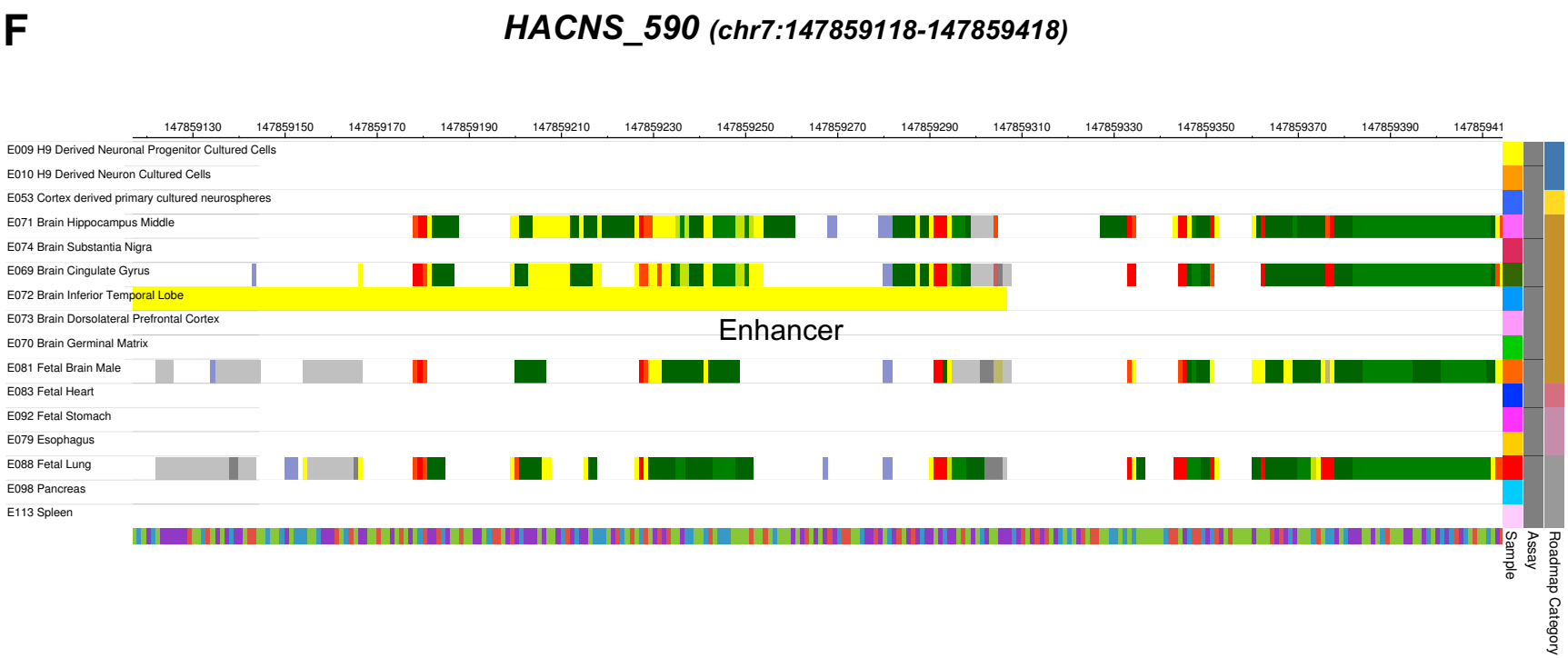


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

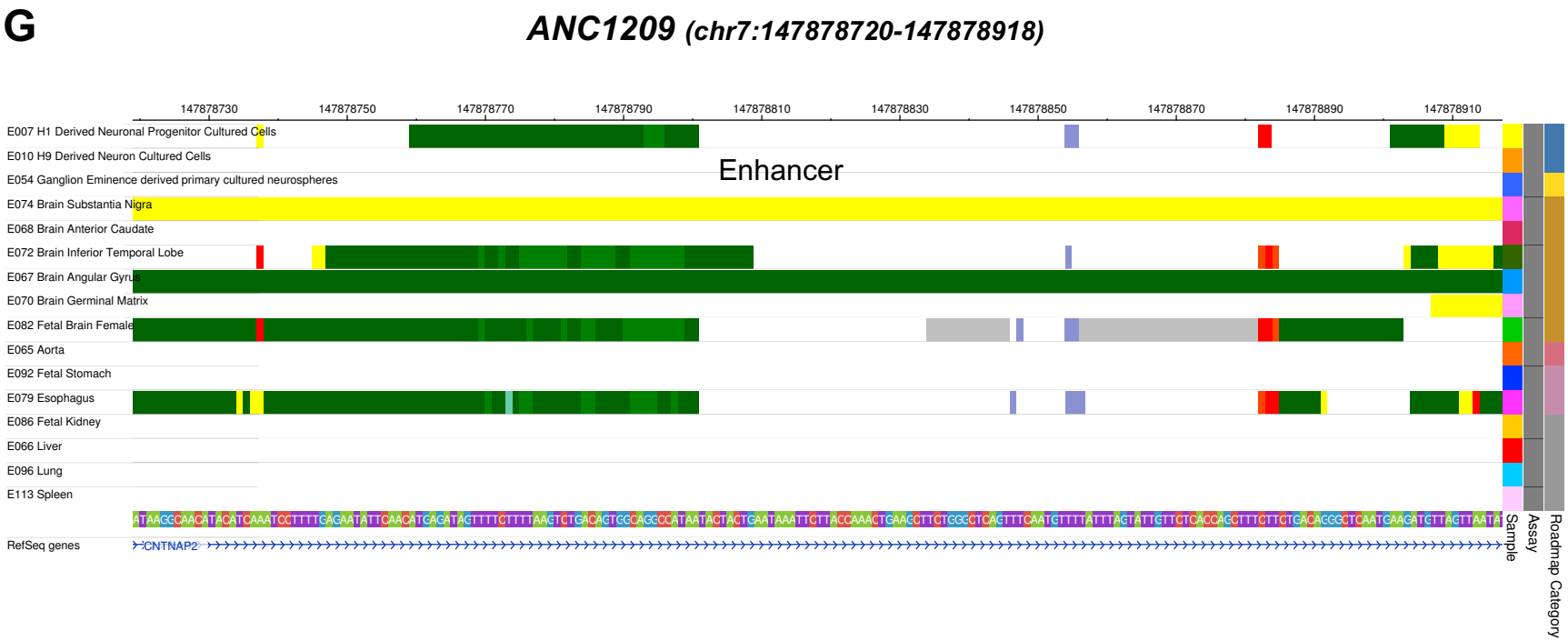


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

H ***HACNS_954 (chr7:148173396-148173905)***

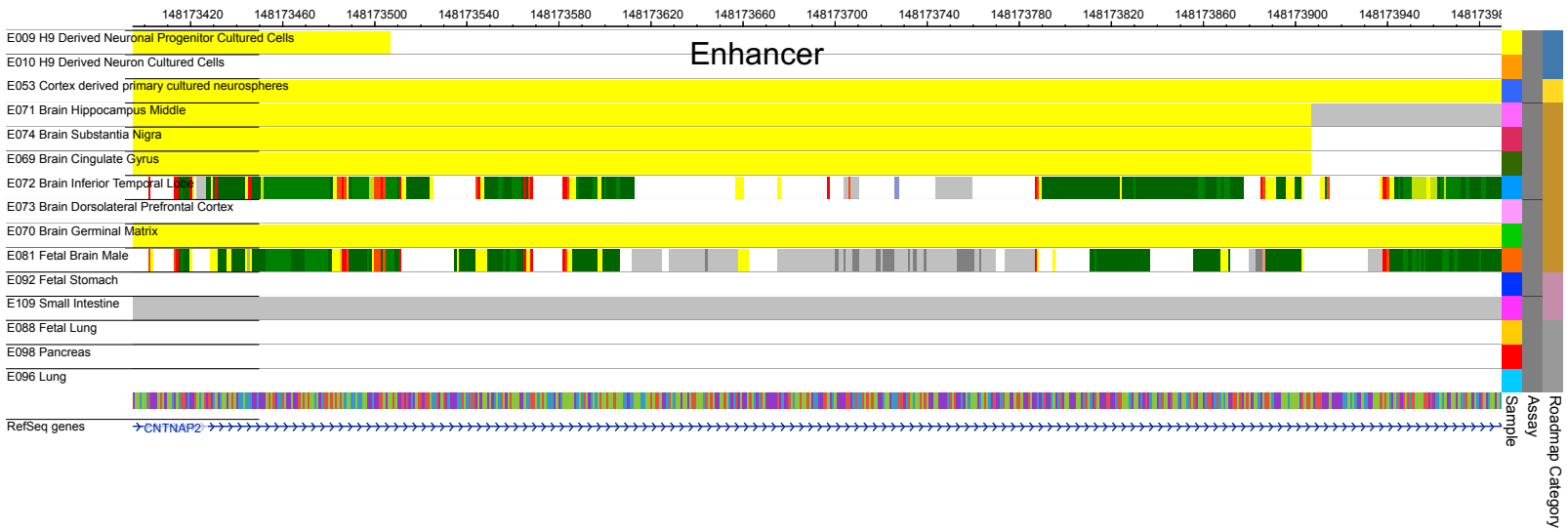


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I. *Figure continued on the next page.*

I

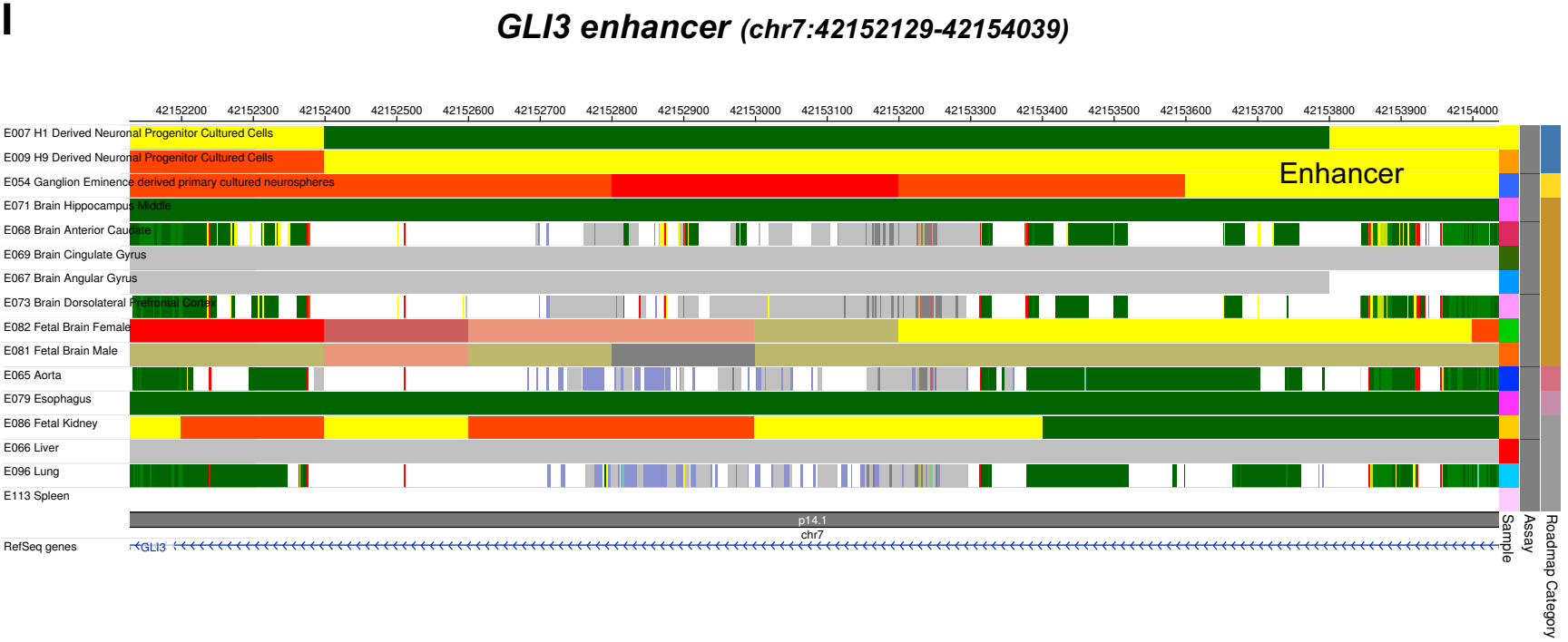


Figure 7.04. ChromHMM annotations for the eight *CNTNAP2* HARs

ChromHMM was used to assign chromatin states to the *CNTNAP2* HARs in primary brain and control epigenomes. Four of the eight HARs were classified as an 'enhancer' (yellow) in one or more of the queried primary brain tissues: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres). The remaining HARs had annotations that included quiescent, weak and strong transcription, heterochromatin, weak repressed polycomb, and transcriptional start sites. Each HAR extends across the full width of the window shown (except for 2xHAR.395 which is only 23bp and is highlighted in yellow). The ChromHMM annotations of a validated forebrain enhancer is shown for context in panel I.

7.7 Luciferase enhancer assay

7.7.1 Results of the luciferase assay

In light of the ChromHMM annotations suggesting some of the *CNTNAP2* HARs may be enhancers, I next wanted to experimentally test their enhancer potential. In order to do so, I tested six of the HARs with a **luciferase reporter assay** (389, 412). ANC1208 and ANC1209 were omitted due to technical issues[10].

Each of the HARs were cloned into a **'pGL3 Promoter' vector** (Promega, E1761A). This vector contains a **SV40 promoter** and **firefly luciferase cDNA** (see **Figure 7.05**). If a HAR is an enhancer, its presence in the vector should increase the luciferase expression. All measurements were performed relative to a baseline vector (the pGL3 Promoter vector alone). To control for differences in transfection efficiency, a second SV40-driven vector was co-transfected with each of the pGL3 promoter plasmids. This **normalization vector**, **'pRL'** (Promega, E2231A), contains cDNA for sea pansy luciferase - which is activated under distinct biochemical conditions from firefly luciferase.

The experimental and normalization constructs were transfected into **SH-SY5Y cells**, an immortalized neuroblastoma cell line. SH-SY5Y cells were used instead of cortical neurons, because post-mitotic neurons have very poor transfection efficiency. Instead of getting low transfection rates – and therefore untrustworthy luciferase readings – I decided to compromise with a non-cortical but neuron-like cell type that would provide more consistent results. As SH-SY5Y cells are a cancerous cell line, any results obtained from the assay would need to be taken with caution. I planned to use the findings from this experiment to prioritize HARs for a more rigorous CRISPR-Cas9

[10] These sequences were difficult to PCR amplify from genomic DNA. Troubleshooting was ongoing at the time of the COVID-19 shutdown.

enhancer assay (see **Chapter 8**). There would therefore be a second round of (more sensitive) experiments to interrogate results from this first assay.

Two independent experiments were conducted, with each construct tested in duplicate per experiment. All assays were performed with the **Dual Luciferase Reporter Assay kit** from Promega (E1910) (see **Figure 7.05** for an overview, and **Chapter 3** for full details). All luciferase recordings were taken on a **luminometer**.

Following data acquisition, Welch's one-way ANOVA was used to test for significant differences in luciferase signals between baseline and HAR-containing plasmids (**Figure 7.06a**). Notably, the ANOVA detected a significant difference amongst the six HARs and baseline luciferase readings [$F(6, 8.87) = 27.4$, $p = 0.00003$]. Tukey tests subsequently identified HACNS_97 as having a significantly increased luciferase signal relative to baseline ($p = 1.51e\text{-}04$). The luciferase readings from HACNS_97 were approximately 2.25-fold higher than baseline (95% confidence interval HACNS_97 = 0.315). All other HARs returned non-significant *p*-values.

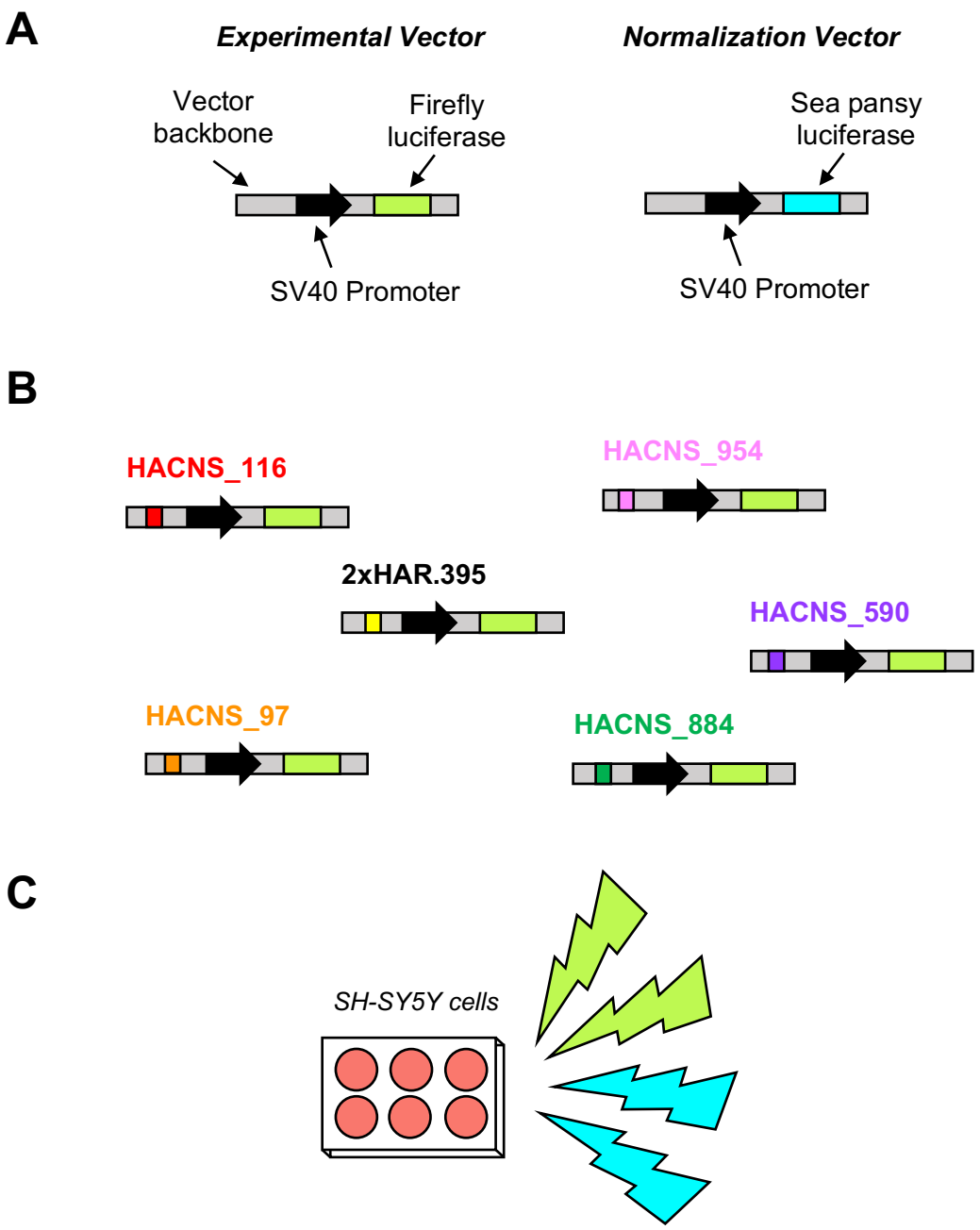


Figure 7.05. Overview of the luciferase enhancer assay

(A) The assay consists of two vectors: 1) an experimental vector that contains a firefly luciferase gene; 2) a normalization vector that contains a sea pansy luciferase gene. The normalization vector is used to control for differences in transfection efficiency between samples. (B) Six of the *CNTNAP2* HARs were cloned into the experimental vector. (C) Each of the modified experimental vectors are co-transfected with the normalization vector into SH-SHY5Y cells (a neuroblastoma cell line). If a HAR possesses enhancer function, the normalized luciferase signal will increase relative to baseline (the firefly vector alone). All luciferase measurements were taken with a luminometer (not shown).

7.7.2 Results of the chimpanzee luciferase assay

Having observed that HACNS_97 showed potential enhancer activity, I next wanted to examine whether similar trends were noted with the chimpanzee HAR orthologues. Crucially, this would mainly serve to clarify whether there was a complete gain or loss of enhancer activity in either humans or chimpanzees. In my experience, luciferase assays are not sensitive enough to robustly detect differences in enhancer strength. Rather, they should be used to observe general trends in a non-quantitative manner. Nevertheless, performing an enhancer assay on the chimpanzee HAR orthologues would be an important test to carry out (in the off chance there is a complete gain or loss).

The chimpanzee luciferase assay was performed in the same manner as described for the human HARs. The chimpanzee sequences were cloned from genomic DNA kindly donated by Alessio Strano. Importantly, due to practical constraints, the chimpanzee and human enhancer assays were performed separately.

Figure 7.06b shows the normalized luciferase readings for the orthologues of each HAR. A one-way ANOVA found no significant difference between luciferase readings for chimpanzee, unlike with the human experiment [$F(6, 19) = 2.18$, $p = 0.407$]. While this result could be taken to suggest a gain of enhancer activity in human HACNS_97, there is clearly a significant amount of noise in our readings, as evident by the large error bars in the plots. The general trend of the data is also quite similar to the human experiment – albeit with even larger numbers. Again, since these two experiments were run separately, with different batches of reagents, conclusions about the relative strength of human versus chimpanzee HAR enhancers should be held off. This is reinforced by the finding that none of the chimpanzee orthologues were significantly different from baseline, despite the larger luciferase readings.

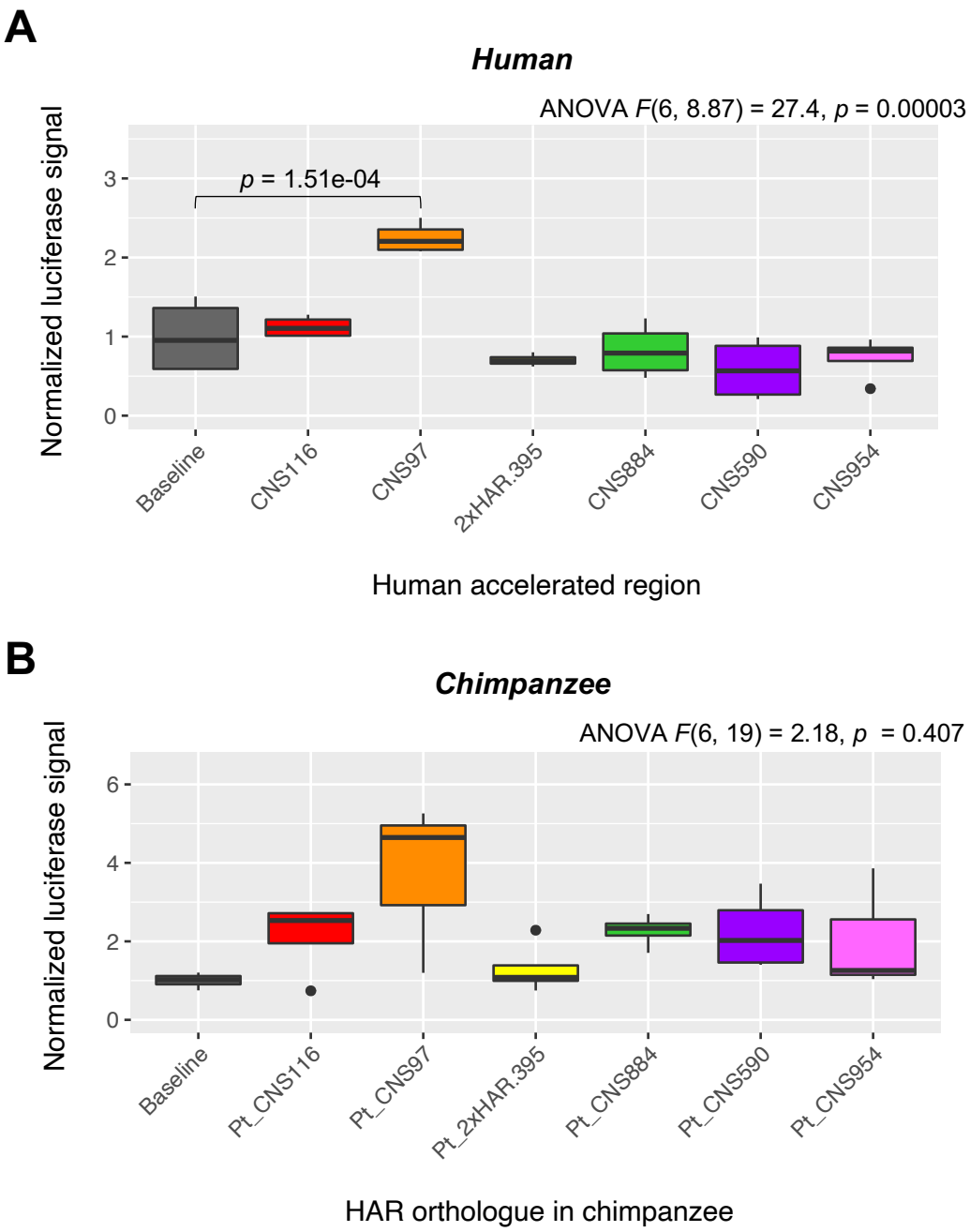


Figure 7.06. Results of the luciferase enhancer assay

(A) Boxplot of the luciferase signals recorded for each HAR across two experiments (n = 2 per HAR for each experiment). Luciferase signals are reported relative to the baseline mean (the experimental vector without a HAR). Results of a one-way ANOVA are reported at the top of the plot. Post-hoc pairwise comparison identified HACNS_97 as having a significantly increased luciferase signal relative to baseline. (B) Luciferase signals from vectors containing the chimpanzee orthologues of each HAR (n = 2 per HAR, repeated for two experiments). No significant difference was observed between any of the HARs and baseline.

7.8 Discussion

In this chapter, I evaluated the genetic variation in human populations in the *CNTNAP2* locus for signatures of positive, negative, and balancing selection. I then followed up these analyses by applying the ChromHMM algorithm to examine the chromatin states of each of the eight *CNTNAP2* HARs. Finally, I experimentally tested the enhancer potential of the HARs using a luciferase enhancer assay.

Firstly, I used the **Tajima's D test** to investigate whether there was evidence for positive or balancing selection at the *CNTNAP2* locus (indicated by negative or positive Tajima's D scores, respectively) (**Figure 7.02** and **Table 7.03**). With respect to **positive selection**, three intron 1 HARs were located in windows that were ranked in the **top 5% most negative** of chromosome 7 (HACNS_116, HACNS_97, and 2xHAR.395). Although it was exciting to find hints of positive selection overlapping some of the HARs, these signatures were population-specific. There was no evidence of species-wide positive selection at any of the HARs (at least using Tajima's D, with the EGDP dataset, and a 2 kb window size). This interpretation is also reinforced by the lack of significantly negative scores in the African genomes (which, as stated, can be used a species-wide proxy). Whether the population-specific negative scores correspond to positive selection or demographic processes (i.e. a population expansion can mimic the signature of reduced genetic variation (63)), will need to be clarified with further neutrality tests and functional experiments.

When considering **positive Tajima's D scores**, ANC1208 and ANC1209 yielded repeatedly significant results. ANC1209, in particular, ranked in the **top 1% most positive** genomic regions in eight of the twelve populations [CSI: top 41 bins, ENE: top 85 bins, SOA: top 71 bins, SSI: top 30 bins, SWE: top 53 bins, VOL: top 121 bins, WAA: top 5 bins, and WSI: top 588 bins]. Of the remaining four populations, it was in the top 5% most positive in three regional populations (AFR, SEA, and SEM). ANC1208 was not quite as remarkable, but was still in the top 1% chromosome-wide for three

populations and in the top 5% in four populations [**top 1%:** NSI: top 638 bins, SEA: top 754 bins, and WSI: top 271 bins; **top 5%:** CSI, ENE, SEM, and WAA].

Finding potential signatures of balancing selection was unexpected, given that the *CNTNAP2* HARs are enriched for nucleotide changes that are fixed in all human populations. An important next step will to be evaluate the significance of the ANC1208 and ANC1209 Tajima's D scores genome-wide (and not just chromosome-wide). Apart from this, there are a number of other reasons why the scores could be significantly positive. As a reminder, the mechanisms that could lead to signals of acceleration were: 1) positive selection, 2) GC-biased gene conversion (gBGC), and/or 3) a loss of constraint (12). In theory, either of the latter two mechanisms could also cause a positive Tajima's D score. That said, GC-biased gene conversion was not detected at either HAR by their associated authors (Lindblad-Toh et al. (10)) or using the phastBias tool (102), an algorithm that predicts gBGC events in the human genome. However, a loss of constraint is certainly a possibility.

A last reason to consider relates to the demographic explanation of a positive Tajima's D score: a population substructure, admixture or contraction event (63). However, considering that positive scores for ANC1208 and ANC1209 were detected in the majority of the 12 populations, this would seem unlikely. The North Siberian population (NSI) was the only group to not have a top 1% or top 5% significant score in ANC1209. This could relate to the size of the NSI population – smaller populations are more vulnerable to genetic drift, which could have removed the signal observed in the other 12 populations (283). Ultimately, functional studies will be needed to clarify whether ANC1208 and ANC1209 were subject to balancing selection or a loss of constraint.

With respect to the *CNTNAP2* gene as a whole, across the 12 populations, **positive Tajima's D scores** clustered in introns 1, 11, 13, and 20. **Negative D values** clustered in introns 1 and 13 (**Figure 7.02**). Given that i) previous selection studies identified positive selection in introns 1 and 13 (13, 14), ii) introns 1 and 13 overlap

disease-associated mutations (**Figures 2.06-2.08, Table 2.02**), and iii) that introns 1 and 13 both contain more than one HAR, these findings are noteworthy. Again, more work will be needed to interrogate their genuine importance for evolution and disease. In a 2.3 Mb gene, however, having preliminary data to help prioritize regions for further study is crucial. As such, these findings would suggest introns 1 and 13, and/or HACNS_116, HACNS_97, ANC1208, and ANC1209 as regions for further investigational focus.

The second neutrality test I performed was the **iHS test**, which detects more recent selection (< 30,000 years ago versus < 250,000 year ago for Tajima's D). As shown in **Table 7.04**, none of the 200 kb windows overlapping *CNTNAP2* were in the top 5% most positive or negative **genome-wide**. That said, there were two windows that ranked in the top 10% genome-wide. Chr7_145.8 Mb ranked in the top 6.8% in Africa, top 8.8% in South/West Europe, and top 7.4% in Volga/Ural. Chr7_146.4 Mb reached the top 8.9% in South Asia and top 5.8% in West Asia/Armenia. The first of the two windows, Chr7_145.8 Mb, contains HACNS_116 and HACNS_97. The second window did not overlap any of the CNTNAP2 HARs. There is therefore not overwhelmingly strong evidence for recent human-wide positive selection at *CNTNAP2*.

In summary, neither the Tajima's D test nor the iHS statistic identified significant signals of human-wide positive selection at either the *CNTNAP2* locus or its eight HARs. There are a number of reasons for why this could be. Firstly, these two tests detected signatures of selection starting from ~250,000 and ~30,000 years ago. It is possible, therefore, that positive selection older than 250,000 years ago could have occurred, but was not detected. There have been an estimated 5-13 million years since humans diverged from chimpanzees (45), the most recent ancestor used by the acceleration studies that identified the *CNTNAP2* HARs. In other words, there was a significant amount of time before 250,000 years ago for positive selection to have fixed the human-specific nucleotide changes. This explanation would also agree with

the finding that almost all of the *CNTNAP2* human-specific nucleotides are shared with Neanderthals/Denisovans (see **Table 2.03**).

Another explanation is that these HARs were not formed by positive selection, but were accelerated by a loss of constraint (as previously discussed). However, it is important to point out that only two HARs were enriched for positive Tajima's D values. This suggests that most of the *CNTNAP2* HARs are unlikely to be caused a loss of constraint (agreeing with the general predictions by Kostka et al. (169)). Similarly, gBGC was also not detected at any of the eight HARs by the phastBias algorithm (102). Again, the most conclusive method of determining whether the HARs are functional/positively selected for is to experimentally test them. While I presented some preliminary functional evidence in this chapter (to be discussed shortly), a more rigorous experiment will be described in the final chapter.

The third study discussed in this chapter was our use of the CADD algorithm to look for negative selection/functional constraint at the *CNTNAP2* HARs (**Figure 7.03**). Overall, six of the eight HARs overlapped local CADD score peaks. HACNS_97 had the highest individual CADD score (22.1), HACNS_590 had the highest mean score (15.5), and both appeared enriched for variants with CADD scores in the top 1% genome-wide – 12.2% and 12.7% of all possible mutations, respectively (**Table 7.05**). These results were notably higher than the mean and minimum scores for chromosome 7 (4.31 and 0.001).

When comparing these CADD scores to the scores obtained for a validated forebrain enhancer in the *GLI3* gene, the HAR CADD scores seem less impressive. The *GLI3* enhancer had strikingly high CADD statistics, with 42% of all variants in the region scoring over 20 (~808/1911 variants). The maximum score – at 24.3 - was also higher than detected in any of the *CNTNAP2* HARs. From this point of view, the *CNTNAP2* HARs may not be as functional or as significant as the *GLI3* enhancer. However, since the CADD algorithm makes use of existing experimental data, and given that the *GLI3* enhancer is listed on databases of human enhancer sequences, it

may not be a fair comparison to make. If the existing functional evidence for the *GLI3* enhancer was incorporated into the CADD algorithm, then the computed CADD scores would be predicted to be high. Since no functional evidence currently exists for the *CNTNAP2* HARs (apart from the DNase I hypersensitive sites (DHSs) previously mentioned), its CADD scores would be expected to be lower. For comparison, I also queried the CADD scores associated with the HARE5 sequence (see **Appendix**). HARE5 is a human accelerated region that was shown to have enhancer activity in a lacZ transgenic mouse assay (172). However, the computed scores/overall trend were quite similar to the *GLI3* enhancer. As with the tests for positive and balancing selection, the only way to definitively test whether purifying selection is at hand is to functionally test the HARs. Like with the Tajima's D and iHS tests, however, the CADD scores point to HACNS_97 as a potentially interesting candidate worthy of prioritization.

Following the computation of CADD scores, I next analyzed each of the HARs for epigenetic marks of enhancer activity by using the ChromHMM algorithm (**Figure 7.04**). Four of the eight HARs were classified as an 'enhancer' in primary brain tissue: 1) HACNS_884 in substantia nigra, 2) HACNS_590 in inferior temporal lobe, 3) ANC1209 in substantia nigra, and 4) HACNS_954 in five tissues (middle hippocampus, substantia nigra, cingulate gyrus, fetal germinal matrix, and cortex-derived primary cultured neurospheres).

As stated at the start of the chapter, several of the *CNTNAP2* HARs were previously shown to overlap DNase I hypersensitive sites (DHSs). HACNS_884 was shown by Gittelman et al. (275) to overlap a human-specific DHS. Won et al. (175) further identified six of the eight HARs as overlapping DHSs in fetal brain (all except 2xHAR.395 and ANC1208). Finally, Capra et al. (170) detected HACNS_884 and HACNS_954 as putative enhancers using their enhancer finding pipeline. They also bioinformatically predicted that HACNS_884 is active in fetal brain, but could not provide a prediction for HACNS_954. With this data in mind, it was highly interesting to note that HACNS_884 and HACNS_954 were also detected as brain-specific

enhancers by ChromHMM. It is curious, however, that only HACNS_954 was detected as an enhancer in fetal brain by ChromHMM, especially given that Won et al. (175) identified most of the *CNTNAP2* HARs as having DHSs in fetal cortex. This discrepancy could be explained by the limited sample sizes used by each respective study, or by the fact that DHSs do not definitively mean a sequence is functional or an enhancer (201).

It is also interesting to consider the ChromHMM results in combination with the findings of our luciferase assay. Only HACNS_97 was detected to increase the normalized luciferase reading relative to baseline (**Figure 7.06**). Given the HARs are placed in an artificial context – outside of their real biological location, in SH-SY5Y cells, in front of a firefly gene – this experiment cannot rule out the possibility that other HARs could be enhancers. Rather, it should be used as an additional line of evidence that HACNS_97 is a HAR to consider for future experimental work.

Taken altogether, the results of this chapter point to the following conclusions. 1) There is no clear-cut evidence for species-wide positive selection at any of the *CNTNAP2* HARs, or in the *CNTNAP2* locus according to the Tajima's D and iHS tests. 2) There are suggestions of potential purifying selection at most of the *CNTNAP2* HARs, according to the CADD algorithm. When compared to validated enhancers, however, the CADD scores of the *CNTNAP2* HARs appear less striking. 3) ChromHMM predicts a HACNS_884, HACNS_954, ANC1209, and HACNS_590 to be enhancers active in the adult brain (or fetal in the case of HACNS_954. 4) A luciferase enhancer assay showed HACNS_97 to have potential enhancer activity, but detected no such effect for any of the remaining HARs tested.

As such, the function and evolutionary importance of these HARs is still somewhat unclear. The experiments conducted in this chapter were important first steps, however, in the absence of clear conclusions, more work is needed. Additional neutrality tests could be performed on the HARs, targeting different timescales and analyzing a larger number of genomes. Additional epigenetic tests, like HiC, could be

employed to determine if there is a physical connection between any of the HARs and another (potential target) locus. Of the limited HiC data that I examined, no interaction was detected between any of the HARs and another locus (2, 275, 413). Most importantly, as repeatedly outlined in this thesis, more functional experiments are needed. These provide the best evidence of whether the *CNTNAP2* HARs (or the gene itself) are functional and important for human-specific cognitive processes in the brain. Detailed next steps, as well as a full synthesis of the thesis' conclusions, will be discussed in the final chapter.

Chapter 8

Conclusions and future work

8.1 Conclusions

Although humans share ~99% genetic identity with chimpanzees and bonobos, our brains differ dramatically in size, structure, and function (1). One key difference pertains to the 'neurites' on human neurons, which are longer and more elaborately branched in the human cortex than in any other primate. This difference is theorized to allow greater information processing in the human brain, and could be a mechanism underpinning human-specialized cognitive functions. Until recently, research into the genetic basis of human brain evolution has been limited. This is partly attributable to the functional and structural complexity of the brain, but also to the overwhelming size of the primate genome (approximately 3 billion nucleotides in size) (414). Over the past decade, however, these roadblocks have been substantially diminished by innovations in two key scientific areas: **i) stem cell neurobiology** and **ii) primate genomics**.

The discovery of methods to generate cortical neurons from pluripotent stem cells ('PSCs'), have allowed us to model differences in human and non-human primate neurons, and crucially, to experimentally study these differences in biologically relevant cells (4, 173). Parallel advances in genomics have accelerated gains in our understanding of the functions of both coding and non-coding DNA. These tools have contributed to several important insights. Firstly, it has become clear that the protein-coding regions of the genome are substantially identical between humans and other primates (415). Moreover, where there are differences, they are usually subtle and unlikely to completely account for human-specific traits (170).

An alternate explanation is that changes to the temporal and spatial patterns of gene expression, rather than to the sequence of proteins they encode, have been the stronger drivers of human evolution. Fuelling this theory was the discovery of ~3000 non-coding DNA sequences termed 'Human Accelerated Regions' (HARs) (9-11, 13, 14, 275). HARs are highly conserved across other mammals, but contain an unusually high number of nucleotide substitutions on the human lineage. Functional genomics data, in combination with machine learning algorithms, have shown that 60% of non-coding HARs are predicted to be gene enhancers (170). Half of these are predicted to target genes active during development, and one third are thought to act in the brain.

Finally, it has become apparent that mutations in genes that differ between humans and other primates often cause severe neurodevelopmental disorders (416). These include autism spectrum disorder (ASD), specific language impairment (SLI), schizophrenia, and a host of intellectual disability (ID) syndromes. Taken together, these observations raised the testable hypothesis that brain-expressed genes which contain HARs may play a key role in human brain development, and that mutations in these genes may cause neurodevelopmental disorders.

Based upon these observations, I focused my PhD on this interface between neurodevelopmental disease research and brain evolution studies. Specifically, I

conducted experiments to begin to dissect out: i) the precise role of 'Contactin-associated Protein-like 2' *(CNTNAP2)* in human brain evolution; and ii) how mutations in *CNTNAP2* cause human-specific diseases like ASD, SLI, and ID. Using a combination of computational and experimental approaches, I aimed to answer the following two key questions:

3) *How do changes to **CNTNAP2 expression** in human cortical neurons affect the development of **neurites** and **dendritic spines**?*
    a. *Does this affect synaptic and network function?*

4) *What are the **functions** of the **HARs** within CNTNAP2?*
    a. *Are they **enhancers** (of CNTNAP2)?*

Answers to these questions will help shed light on what makes us human. As stated throughout the thesis, they may also inform on the causes and potential treatments of diseases associated with *CNTNAP2* (22). *CNTNAP2* was selected as a candidate based on the presence of eight human accelerated regions (HARs) within the gene, its association with cognitive disorders, and previous work in mice showing *CNTNAP2* KO reduces neurite branching and dendritic spine density. Immediate next steps will be addressed over the subsequent pages. This will be followed by a discussion of the main findings of the thesis, and their implications for the fields of human evolution and disease.

8.2 Thesis Discussion

With respect to the first thesis aim - *how do changes to CNTNAP2 expression affect the development of neurites and dendritic spines in human cortical neurons?* – the work presented in this thesis points to a number of conclusions. In **Chapter 6**, I showed that human *CNTNAP2* KO neurons develop significantly fewer neurites, most likely dendrites based on their calibre, than their WT counterparts. This phenotype was observed from D55-D66 (when the experiment ended), and only when WT and KO cells were plated in the same wells. While no decrease in dendritic spine density was observed, given that the experiment only contained a small sample size, I will refrain from discussing the effect of *CNTNAP2* on dendritic spines in great detail.

Previous studies have suggested a few mechanisms for the KO-associated reduction in dendritic branching. These mechanisms ultimately cause either 1) reduced neurite outgrowth or 2) reduced neurite stability. There is currently more evidence for altered neurite outgrowth. Specifically, adhesive contacts during axon outgrowth involve cadherins, integrins and members of the **immunoglobulin superfamily (IgSF-CAMs)** (261). As CASPR2 is known to interact with IgSF-CAMs in mouse axons, it is possible that human CASPR2 interacts with IgSF-CAMs to regulate neurite outgrowth via cell adhesion and/or cytoskeleton remodelling at growth cones (193). A lack of CASPR2 could disrupt this process, and thus lead to reduced branching. Additionally, CASPR2 could mediate the binding between **contactins** and **receptor phospho-tyrosine phosphatases** (recall, CASPR2 binds to contactin-2) (35, 417, 418). This binding is known to occur in rodents (35). In turn, this may activate dendritic growth signaling via the protein tyrosine kinase pathway (35, 419). Intriguingly, our scRNA-Seq data identified *DOK5* as down-regulated in KO immature cortical neurons. This finding is highly interesting because *DOK5* is part of the receptor tyrosine kinase pathway (420). Even more striking is that there is existing data in mice showing *DOK5* mediates neurite outgrowth. This therefore provides a second mechanism by which *CNTNAP2* KO could cause reduced neurite branching.

In a separate mouse study, *Cntnap2* KO was shown to reduce interneuron dendritic branching (202). This was due to decreased dendrite stabilization rather than decreased outgrowth. The decreased stabilization was identified to be occurring through the **calcium/calmodulin-dependent serine protein kinase (CASK)**. Loss of *CNTNAP2* resulted in a failure of CASPR2 to bind to CASK, and thus promote dendrite maintenance. If this mechanism is also occurring in human excitatory neurons, *CNTNAP2* KO could lead to a reduction in the number of branches by causing a loss of CASK-mediated neurite stability. The fact that reduced branching was only noted at the final three timepoints in the assay (i.e. and not earlier), would support this hypothesis. Beyond these theories, however, further work will be needed to clarify the precise mechanisms underlying our reduced branching phenotype.

In that same vein, the molecular roots of the increased neuronal activity observed in the *CNTNAP2* KO cells is also unclear. The results of the scRNA-Seq experiment presented in **Chapter 5**, however, provide some suggestions. For example, the increase in *SYNPR* indicates an increased number of synapses present in KO cortical neurons. More synapses could explain the stronger, more frequent, and more synchronized network bursts. To validate this theory, synaptic puncta would need to be quantified with immunostaining. Another explanation (supported by *Cntnap2* KO rodent studies) is altered AMPA receptor trafficking (34, 203, 262). While the rodent studies found loss of *Cntnap2* lead to GluA1 aggregations (and reduced activity), it is possible that *CNTNAP2* KO in humans affects receptor trafficking in another way. For example, by increasing the number of receptors and thus causing an increase in neuronal excitation. Alternatively, Scott et al. (194) detected a significant increase in neurotransmitter release in their *Cntnap2* KO mouse model. This is another mechanism that could cause the over-excitation I observed. A last possibility relates to the finding that *Cntnap2* KO in mice reduces the number of voltage gated potassium channels (193, 194). Since these channels are needed to re-polarize axons after an action potential, it could be that a loss of these channels makes the *CNTNAP2* KO cells less polarized, and therefore easier to excite.

Overall, these findings are highly useful for furthering our understanding of human brain evolution and disease. Deciphering the cellular/molecular processes that get disrupted in *CNTNAP2* mutations is necessary to treat the patients harbouring them. The work I present in this thesis will help achieve this goal in a few key ways. As previously stated, I provide the second ever human *CNTNAP2* KO cortical line (to my knowledge). I used this line to conduct the first studies of how *CNTNAP2* affects the development and function of neurites, dendritic spines, and neuronal activity. This model indicated that over-excitation and altered neuronal structure (possibly leading to a reduction in connectivity) are likely mechanisms accounting for *CNTNAP2* disease phenotypes (e.g. epilepsy, autism, intellectual disability). Hopefully, our model will be used by other scientists to investigate additional aspects of *CNTNAP2* function/dysfunction. The differing results in our human model compared to those published in rodent studies, further supports that human models are important for understanding human phenotypes.

Additionally, our scRNA-Seq dataset provided candidate genes/pathways that may be relevant for *CNTNAP2* disease treatments. Again, to my knowledge, this is the first transcriptomic study of *CNTNAP2* loss-of-function (in any species) conducted at single cell resolution. While further studies will be needed to understand which (if any) of the DE genes and/or vulnerable cell types are vital to *CNTNAP2* disease pathogenesis, our list provides a starting point to build off from (along with three existing bulk RNA-Seq datasets (239, 295, 307)). They also reinforce that alterations to neurotransmission may be crucial components of *CNTNAP2*-associated diseases (as is evident from the DE genes that fell into this category). They therefore offer further suggestion that targeting neurotransmission will be a fruitful strategy for therapeutic investigations.

These findings are also relevant for a possible role of *CNTNAP2* in human evolution. As is known from human disease models, changes to either excitation or inhibition can lead to profound effects on cognition. Increased excitation has been implicated in a variety of neurodevelopmental disorders including Fragile X syndrome

(421-423) and autism (380, 424). Our qRT-PCR and western blot data showed that *CNTNAP2* expression was significantly increased in human cortical neurons relative to macaques. This result falls in line with the theory that changes to the temporospatial expression of genes – and not changes to the proteins they encode – are critical for human-primate differences. It also provides preliminary grounds to suggest the following:

If *CNTNAP2* expression is reduced in non-human primate cortex, and assuming the phenotype caused by a complete loss of *CNTNAP2* would be akin to that caused by a reduction, *CNTNAP2* may be important for differentially regulating neuronal activity between humans and other primates. Although this may seem like a large assumption, the fact that *CNTNAP2* is dosage-sensitive (i.e. heterozygous mutations cause disease), clearly emphasizes how changes to the levels of this gene can produce strong phenotypes. This could cause a significant contribution to the differing cognitive abilities between humans and our closest living relatives. As stated in **Chapter 6**, while direct electrophysiological comparisons of human and primate neural activity are rare, there is some initial data to suggest there are differences in human and primate neural activity (29, 388).

The human increase in *CNTNAP2* cortical expression also connects heavily to the HARs. The second aim of my thesis was to investigate the functions of the HARs within *CNTNAP2*, and particularly, to ask if they are enhancers. Our finding that certain of the *CNTNAP2* HARs overlapped epigenetic marks of enhancers (via ChromHMM), in conjunction with the identification of one HAR, HACNS_97, as an enhancer in a luciferase reporter experiment, suggests the HARs may be an important mechanism leading to the increased expression of *CNTNAP2* in humans. While I was not able to provide concrete evidence showing the HARs enhance *CNTNAP2*, this will be an important next step. The scRNA-Seq approach I described above will go some way to address this and other questions. This finding also agrees with the prediction by Capra et al. (170) that many HARs are enhancers – and more specifically, that a large subset are enhancers active in the developing brain.

While the neutrality tests did not identify signals of human-wide positive selection at the *CNTNAP2* locus, or any of the HARs, this is not entirely unexpected. As described in **Chapter 7**, Tajima's D and the integrated haplotype score only detect selection from ~250 KYA and ~30 KYA, respectively (168). Therefore, for the *CNTNAP2* gene to be of evolutionary importance in humans, does not necessarily mean these two tests should return a significant signal. If selection had acted prior to 250 KYA, then it would not be detected but could still be important for human evolution. It was worth performing these tests, however, given the possibility that selection at *CNTNAP2* lasted until these more recent timescales. It was also important as a means to improve our understand of the gene's relevance in different human populations. The lack of recent positive selection is further explained by the fact that the *CNTNAP2* HARs mostly share their human-specific substitutions with archaic humans (i.e. Neanderthals and Denisovans) (166). This suggests the evolution of these sequences occurred before the Neanderthal/Denisovan – human split, which occurred much earlier than 250 KYA (425). Moreover, in light of the HARs containing many CADD scores >20, this indicates that that there is potential purifying selection. This could be interpreted as further support for the notion of positive selection at the HARs. Specifically, following the completion of positive selection, these loci may then be subject to purifying selection to functionally constrain them.

The unexpected findings from **Chapter 7** were that two of the *CNTNAP2* HARs, ANC1208 and ANC1209, overlapped positive Tajima's D values (significant chromosome-wide). Positive Tajima's D scores can be a sign of balancing selection, population bottlenecks, or potentially a lack of constraint. While it was not possible to disentangle these causes in my thesis, this does point to a few wider points. While HARs are extremely valuable as a means to find **distant positive selection** in **non-coding portions** of the genome, HARs may not always be caused by positive selection. That said, of the eight HARs in *CNTNAP2*, only two showed this signal. Curiously, both were from the same study (Linblad-Toh et al. (10)), highlighting the fact that different acceleration tests may be more accurate at detecting positive selection than others.

However, overall this finding does not detract substantially from the utility of acceleration tests. This is especially true considering that most other tests for distant selection can only be applied on protein-coding DNA (168). Our results simply confirm a caveat to their use that has been previously described by others (12). Acceleration tests are best used as a means to prioritize genomic regions for further study, and not as a guarantee that they were positively selected for.

Until then, the work presented here provides preliminary evidence that *CNTNAP2* contributes to neurodevelopmental diseases through altering neuronal activity. This may be underpinned by changes to neurite branching, and ultimately, to neuronal connectivity. With respect to evolution, apart from the eight HARs within the *CNTNAP2* locus, there is no indication of human-wide positive selection at the gene. However, there is evidence of increased *CNTNAP2* expression in the human cortex relative to other primates. I also provide data that one or more of the *CNTNAP2* HARs are enhancers. As such, these HARs could be contributing to the species-difference in *CNTNAP2* cortical expression, and thus, may be regulating neuronal activity to allow human-specialized brain function.

8.3 Next steps

Before the start of the COVID-19 lockdown, I was working on a **HAR scRNA-Seq** experiment to further test the enhancer potential of the *CNTNAP2* HARs. The premise of this experiment was to use pairs of CRISPR-Cas9 guide RNAs (gRNAs) to excise the *CNTNAP2* HARs in pluripotent stem cells (PSCs) (see **Figure 8.01**). I would then differentiate the edited PSCs (along with the un-edited WT line) to a forebrain identity using the Shi et al. (181) protocol. With these induced cultures, I would use scRNA-Seq to assess whether the loss of a HAR significantly reduced or increased expression of certain genes. A reduction would hint at the HAR being an enhancer of that gene; conversely, an increase would hint it could be a repressor.

scRNA-Seq would be used for three key reasons: 1) it allows a genome-wide search for HAR target genes, 2) it allows cell type-specific effects to be detected, and 3) it would study enhancer potential in a biologically relevant context. The first reason is relevant because it cannot be assumed the *CNTNAP2* HARs are enhancing *CNTNAP2* (provided they are indeed enhancers). Enhancers have been shown to act on target genes millions of base pairs away (172). The second reason is also important because if the HAR is acting as an enhancer in one specific cell type, scRNA-Seq gives the resolution to detect this. If, for example, a HAR was enhancing its target in cortical hem only (i.e. a small population of cells in the culture), the enhancer-target interaction could be diluted out using traditional bulk assays. Finally, the third reason addresses the shortcomings of the luciferase assay described in **Chapter 7**. The enhancer potential of a HAR could be tested in its true genomic location, on its real target genes, and in human forebrain cells.

Testing more independent inductions will be critical to accurately determine whether a HAR is definitely enhancing/repressing any identified genes. In transcriptome-wide experiments, certain genes will be differentially expressed for no biological reason (i.e. due to noise). Therefore, a real target gene should be

differentially expressed across multiple inductions/experiments. There are also other methods of validation, such as chromosome conformation capture, that can test for a physical interaction between HARs and their putative targets (172). Multiple rounds of scRNA-Seq, in addition to qRT-PCR and western blots, should be used to further validate any observations.

As of September 2020, I have successfully deleted two HARs from the NDC1.2 PSC line. These two HARs were HACNS_97 and HACNS_954, located in introns 1 and 18, respectively. They were chosen for a variety of reasons, based on the results of this thesis and based on predicted guide pair efficiencies. HACNS_97 was the only HAR to show enhancer potential in the luciferase assay. It also overlapped potential signatures of positive selection (Tajima's D in SEM and SSI populations, iHS in AFR, SWE, and VOL). Similarly, the CADD algorithm showed it contained one of the highest proportions of variants (out of the eight HARs) in the top 1% most deleterious genome-wide. HACNS_954, conversely, was predicted to be an enhancer in five brain tissues by ChromHMM (cortex derived primary cultured neurospheres, hippocampus, substantia nigra, cingulate gyrus, and fetal germinal matrix). It was also found to overlap DNase I hypersensitive sites (DHSs) in publications by multiple independent groups (Won et al. (175) and Capra et al. (170)). Critically, both HARs had pairs of gRNAs that were 1) located close to the boundaries of the HAR (allowing for a specific deletion), and 2) predicted to have high efficiency and low probabilities of off-target effects. The other six HARs did not have guide pairs as optimal as these two. **Figure 8.02** illustrates the deletion approaches for each HAR, along with sanger sequencing and PCR validations of the deletions (see **Chapter 3** for full details). When I return to the Livesey lab for a post-doctoral position in November 2020, completing this scRNA-Seq experiment will be a priority.

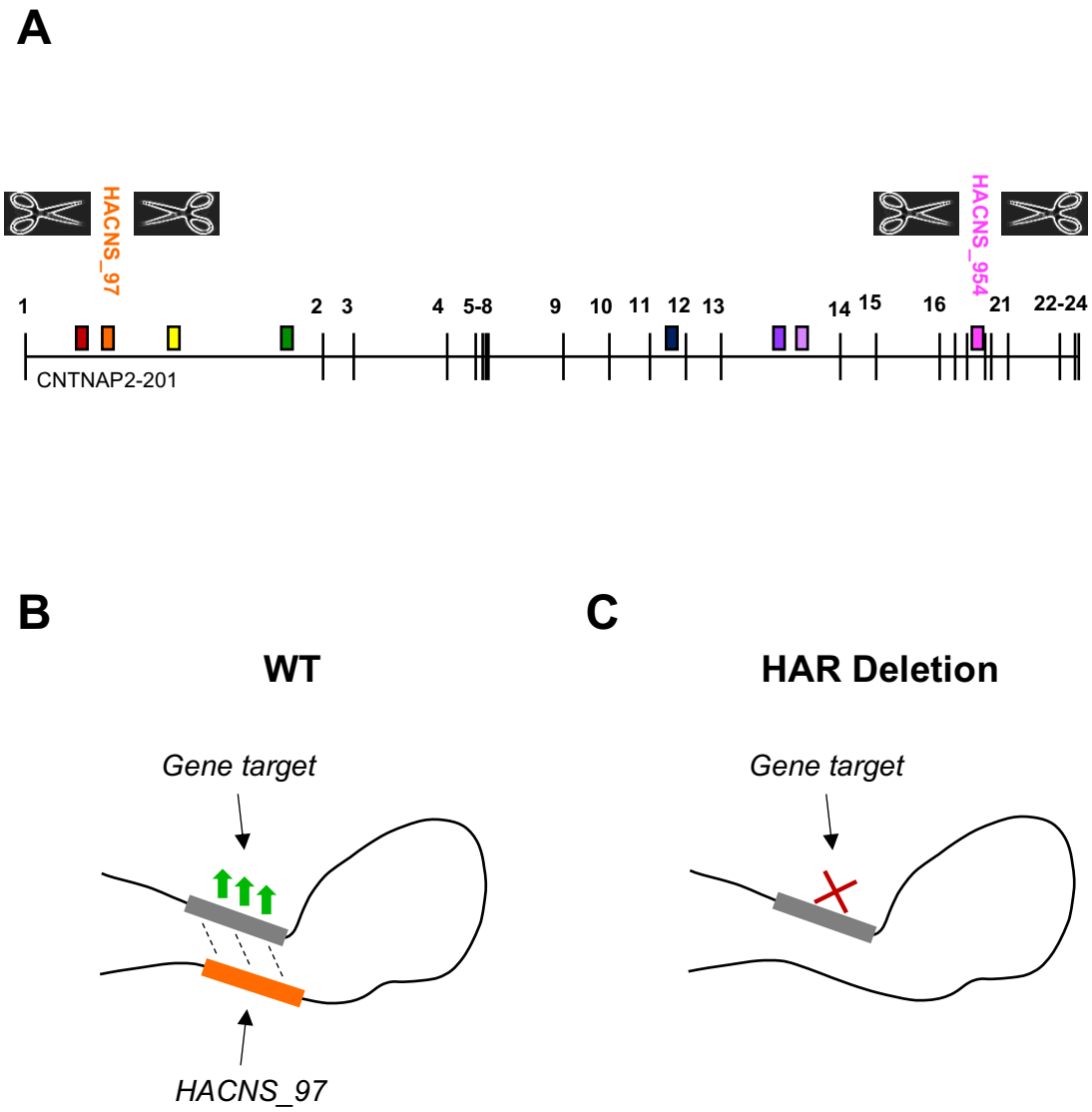


Figure 8.01. HAR single cell RNA sequencing (scRNA-Seq) approach

(A) Pairs of CRISPR-Cas9 guide RNAs (gRNAs) were designed to excise HACNS_97 and HACNS_954 from the genomes of human PSCs. Care was taken to identify gRNA pairs that cut as closely to the HAR boundaries as possible. This would ensure the deletion was specific. (B) The scRNA-Seq approach relies on the principle that if a HAR is an enhancer of a gene, loss of the HAR will cause a reduction in that gene's expression. scRNA-Seq provides an unbiased method to search for HAR targets genome-wide. It also allows cell-type specific enhancer activity to be detected. Finally, it permits HAR enhancer function to be studied in a biologically relevant context (i.e. in human forebrain cells) and without synthetic DNA constructs (i.e. a firefly luciferase vector).

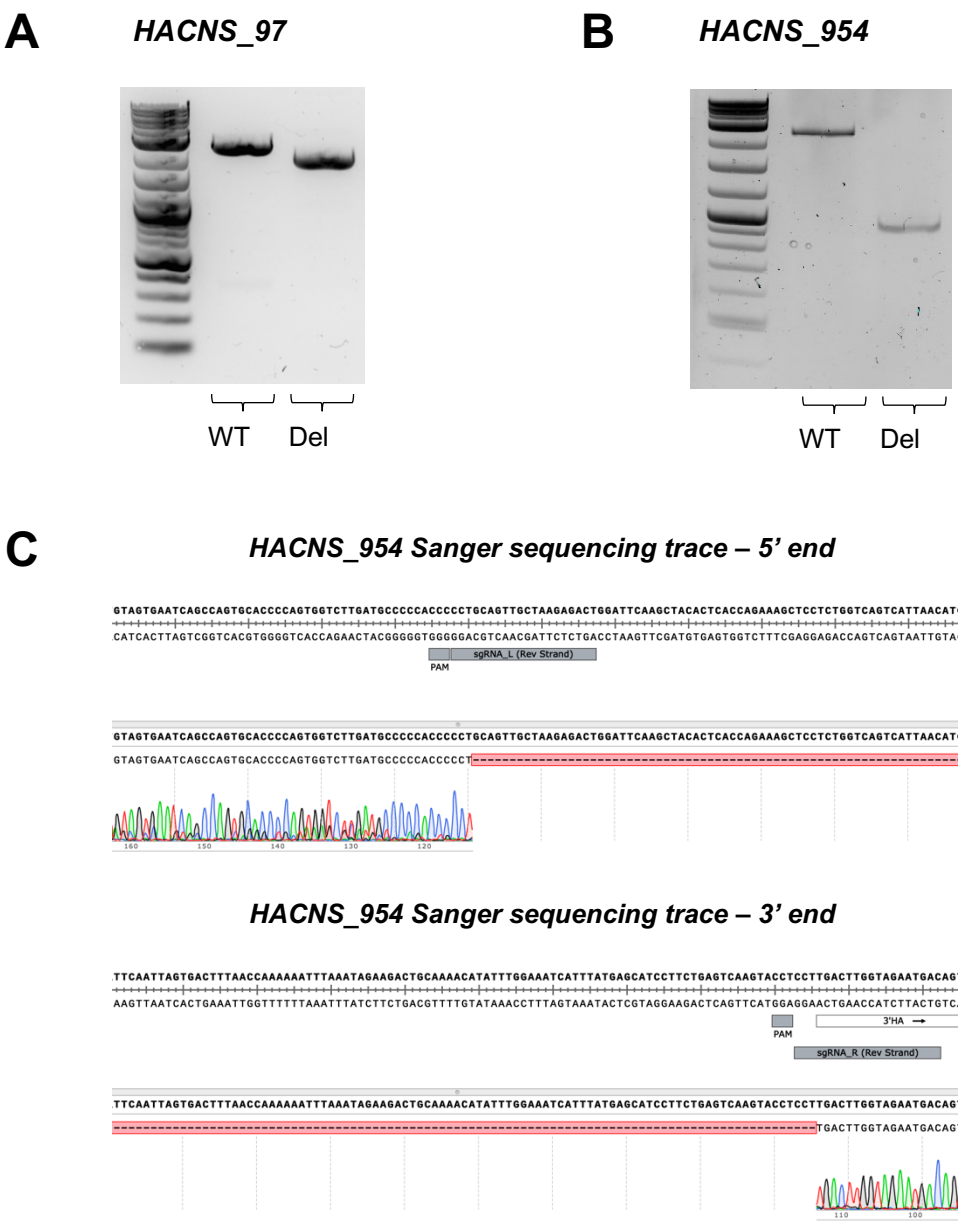


Figure 8.02. CRISPR-Cas9 mediated deletion of HACNS_97 and HACNS_954

Agarose gels showing the PCR products of HACNS_97 and HACNS_954 from genomic DNA. (A) In the HACNS_97 gel, the WT sample shows a clear band at ~2500 bp (the expected size for the amplified region). The HAR deletion sample, however, shows a DNA band at 1900 bp (i.e. minus the expected 600 bp deletion). (B) For the HACNS_954 gel, the WT sample shows a band at ~2300 bp (the expected size), while the HAR deletion sample shows a band at 1000 bp (i.e. minus the expected 1300 bp deletion). (C) Sanger sequencing trace around the HACNS_954 5' gRNA cut site, and another trace for the region containing the 3' gRNA cut site. As shown, no DNA is present past the 5' or 3' cut sites. Together with the PCR results, this shows both HACNS_97 and HACSN_954 have been successfully deleted from the PSC genome. (Sanger sequencing trace for HACNS_97 not shown for space).

As mentioned throughout each results chapter, there are a number of other next steps to take. These included:

i) **Generating additional *CNTNAP2* CRISPR-Cas9 lines.** These included lines with other homozygous loss-of-function mutations, heterozygous mutations, and corrections of patient mutations. It would be highly useful to test whether heterozygous mutations cause intermediate phenotypes (i.e. mutations are dosage sensitive), and why different mutations cause different phenotype severities (43).

ii) **Performing additional scRNA-Seq studies on ventralized cultures.** The absence of cortical interneurons in the inductions we studied with scRNA-Seq (**Chapter 5**) prohibited us from testing whether *CNTNAP2* KO leads to a significant loss of interneurons. Studies from rodent *Cntnap2* KO models suggests > 30% losses occur (245, 269, 271, 272). This is an important phenotype to research, given it could explain the disease phenotypes associated with human *CNTNAP2* mutations (e.g. cortical dysplasia focal epilepsy, Pitt-Hopkins Syndrome, autism, schizophrenia, and others) (43).

iii) **Repeating the neurite length/branching assay at D80.** As mentioned in **Chapter 6**, our assays ended at D66. However, the over-activity phenotype observed in the KO only began appearing around then. It will be important to clarify whether decreased branching in the KO is still noted after D60. This could illuminate whether the increase in activity was caused by greater branching (as shown in the *Cntnap2* KO rodent models (35)), or if another mechanism needs to be investigated.

iv) **Performing additional culture activity assays.** The main short-coming of the Incucyte activity assay was its small sample size (two inductions of WT and two inductions of KO, n = 8 wells per induction). It will be important to see whether the significantly increased activity observed in the KO cells replicates. It will also be necessary to validate these observations using other gold-standard techniques (e.g. patch-clamping or multi-electrode assays).

v) **Performing additional enhancer assays.** The luciferase assay described in **Chapter 7** provided a straight-forward/'first look' approach. However, this experiment was by no means perfect. The experiment was conducted in an SH-SY5Y cell line, using a firefly gene, on a synthetic vector. The scRNA-Seq experiment discussed above will be a useful way to deal with such limitations.

References

1. Sousa AMM, Meyer KA, Santpere G, Gulden FO, Sestan N. Evolution of the Human Nervous System Function, Structure, and Development. Cell. 2017;170(2):226-47.

2. Doan RN, Bae BI, Cubelos B, Chang C, Hossain AA, Al-Saad S, et al. Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. Cell. 2016;167(2):341-54.e12.

3. Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, et al. Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. Cell. 2019;176(4):743-56.e17.

4. Otani T, Marchetto MC, Gage FH, Simons BD, Livesey FJ. 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. Cell Stem Cell. 2016;18(4):467-80.

5. Merkle FT, Eggan K. Modeling human disease with pluripotent stem cells: from genome association to function. Cell Stem Cell. 2013;12(6):656-68.

6. Preuss TM, Caceres M, Oldham MC, Geschwind DH. Human brain evolution: insights from microarrays. Nat Rev Genet. 2004;5(11):850-60.

7. Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, et al. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature. 2009;462(7270):213-7.

8. Konopka G, Friedrich T, Davis-Turak J, Winden K, Oldham MC, Gao F, et al. Human-specific transcriptional networks in the brain. Neuron. 2012;75(4):601-17.

9. Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, et al. Forces shaping the fastest evolving regions in the human genome. PLoS genetics. 2006;2(10):e168.

10. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011;478(7370):476-82.

11. Prabhakar S, Noonan JP, Paabo S, Rubin EM. Accelerated evolution of conserved noncoding sequences in humans. Science. 2006;314(5800):786.

12. Hubisz MJ, Pollard KS. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. Curr Opin Genet Dev. 2014;29:15-21.

13. Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, Beazley C, et al. Fast-evolving noncoding sequences in the human genome. Genome Biol. 2007;8(6):R118.

14. Bush EC, Lahn BT. A genome-wide screen for noncoding elements important in primate evolution. BMC evolutionary biology. 2008;8:17.

15. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, et al. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature. 2005;437(7055):88-93.

16. Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, et al. Gene copy number variation spanning 60 million years of human and primate evolution. Genome Res. 2007;17(9):1266-77.

17. Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, et al. Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol. 2004;2(7):E207.

18. Goidts V, Armengol L, Schempp W, Conroy J, Nowak N, Muller S, et al. Identification of large-scale human-specific copy number differences by inter-species array comparative genomic hybridization. Hum Genet. 2006;119(1-2):185-98.

19. Armengol G, Knuutila S, Lozano JJ, Madrigal I, Caballin MR. Identification of human specific gene duplications relative to other primates by array CGH and quantitative PCR. Genomics. 2010;95(4):203-9.

20. Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, et al. Copy number variation and evolution in humans and chimpanzees. Genome Res. 2008;18(11):1698-710.

21. Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C, et al. Evolution and diversity of copy number variation in the great ape lineage. Genome Res. 2013;23(9):1373-82.

22. Geschwind DH, Rakic P. Cortical evolution: judge the brain by its cover. Neuron. 2013;80(3):633-47.

23. Somel M, Liu X, Khaitovich P. Human brain evolution: transcripts, metabolites and their regulators. Nat Rev Neurosci. 2013;14(2):112-27.

24. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005;437(7055):69-87.

25. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188(4184):107-16.

26. Liu X, Somel M, Tang L, Yan Z, Jiang X, Guo S, et al. Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. Genome Res. 2012;22(4):611-22.

27. Bianchi S, Stimpson CD, Duka T, Larsen MD, Janssen WG, Collins Z, et al. Synaptogenesis and development of pyramidal neuron dendritic morphology in the chimpanzee neocortex resembles humans. Proceedings of the National Academy of Sciences of the United States of America. 2013;110 Suppl 2:10395-401.

28. Bianchi S, Stimpson CD, Bauernfeind AL, Schapiro SJ, Baze WB, McArthur MJ, et al. Dendritic morphology of pyramidal neurons in the chimpanzee neocortex: regional specializations and comparison to humans. Cerebral cortex (New York, NY : 1991). 2013;23(10):2429-36.

29. Marchetto MC, Hrvoj-Mihic B, Kerman BE, Yu DX, Vadodaria KC, Linker SB, et al. Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. Elife. 2019;8.

30. Rakic P. Evolution of the neocortex: a perspective from developmental biology. Nat Rev Neurosci. 2009;10(10):724-35.

31. Defelipe J. The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. Front Neuroanat. 2011;5:29.

32. Benavides-Piccione R, Ballesteros-Yanez I, DeFelipe J, Yuste R. Cortical area and species differences in dendritic spine morphology. J Neurocytol. 2002;31(3-5):337-46.

33. Lima Caldeira G, Peça J, Carvalho AL. New insights on synaptic dysfunction in neuropsychiatric disorders. Current Opinion in Neurobiology. 2019;57:62-70.

34. Varea O, Martin-de-Saavedra MD, Kopeikina KJ, Schurmann B, Fleming HJ, Fawcett-Patel JM, et al. Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. Proc Natl Acad Sci U S A. 2015;112(19):6176-81.

35. Anderson GR, Galfin T, Xu W, Aoto J, Malenka RC, Sudhof TC. Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. Proc Natl Acad Sci U S A. 2012;109(44):18120-5.

36. Gdalyahu A, Lazaro M, Penagarikano O, Golshani P, Trachtenberg JT, Geschwind DH. The Autism Related Protein Contactin-Associated Protein-Like 2 (CNTNAP2) Stabilizes New Spines: An In Vivo Mouse Study. PLoS One. 2015;10(5):e0125633.

37. Nowick K, Gernat T, Almaas E, Stubbs L. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. Proc Natl Acad Sci U S A. 2009;106(52):22358-63.

38. Schneider E, El Hajj N, Richter S, Roche-Santiago J, Nanda I, Schempp W, et al. Widespread differences in cortex DNA methylation of the "language gene" CNTNAP2 between humans and chimpanzees. Epigenetics. 2014;9(4):533-45.

39. Muntane G, Horvath JE, Hof PR, Ely JJ, Hopkins WD, Raghanti MA, et al. Analysis of synaptic gene expression in the neocortex of primates reveals evolutionary changes in glutamatergic neurotransmission. Cereb Cortex. 2015;25(6):1596-607.

40. Ayub Q, Yngvadottir B, Chen Y, Xue Y, Hu M, Vernes SC, et al. FOXP2 targets show evidence of positive selection in European populations. Am J Hum Genet. 2013;92(5):696-706.

41. Mozzi A, Forni D, Clerici M, Pozzoli U, Mascheretti S, Guerini FR, et al. The evolutionary history of genes involved in spoken and written language: beyond FOXP2. Sci Rep. 2016;6:22157.

42. Kamm GB, Pisciottano F, Kliger R, Franchini LF. The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. Mol Biol Evol. 2013;30(5):1088-102.

43. Rodenas-Cuadrado P, Ho J, Vernes SC. Shining a light on CNTNAP2: complex functions to complex disorders. Eur J Hum Genet. 2014;22(2):171-8.

44. Poot M. Connecting the CNTNAP2 Networks with Neurodevelopmental Disorders. Molecular syndromology. 2015;6(1):7-22.

45. Kuhlwilm M, de Manuel M, Nater A, Greminger MP, Krützen M, Marques-Bonet T. Evolution and demography of the great apes. Curr Opin Genet Dev. 2016;41:124-9.

46. van Holstein L, Foley RA. Hominin Evolution. In: Shackelford TK, Weekes-Shackelford VA, editors. Encyclopedia of Evolutionary Psychological Science. Cham: Springer International Publishing; 2017. p. 1-22.

47. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, et al. Insights into hominid evolution from the gorilla genome sequence. Nature. 2012;483(7388):169-75.

48. Lankau EW, Turner PV, Mullan RJ, Galland GG. Use of nonhuman primates in research in North America. J Am Assoc Lab Anim Sci. 2014;53(3):278-82.

49. Rogers J, Gibbs RA. Comparative primate genomics: emerging patterns of genome content and dynamics. Nature reviews Genetics. 2014;15(5):347-59.

50. Boyle E, Wood B. Human Evolutionary History. In: Kaas J, editor. *Evolution of Nervous Systems Volume 4 The Evolution of the Human Brain: Apes and other Ancestors*2017.

51. Bradley BJ. Reconstructing phylogenies and phenotypes: a molecular view of human evolution. J Anat. 2008;212(4):337-53.

52. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. Nature. 2013;499(7459):471-5.

53. Carroll SB. Genetics and the making of Homo sapiens. Nature. 2003;422(6934):849-57.

54. Lehman SM. Introduction to evolutionary anthropology. Toronto: Pearson Canada; 2010.

55. Villmoare B, Kimbel WH, Seyoum C, Campisano CJ, DiMaggio EN, Rowan J, et al. Paleoanthropology. Early Homo at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia. Science (New York, NY). 2015;347(6228):1352-5.

56. Antón SC. Natural history of Homo erectus. Am J Phys Anthropol. 2003;Suppl 37:126-70.

57. Rightmire GP. Out of Africa: modern human origins special feature: middle and later Pleistocene hominins in Africa and Southwest Asia. Proc Natl Acad Sci U S A. 2009;106(38):16046-50.

58. Roebroeks W, Villa P. On the earliest evidence for habitual use of fire in Europe. Proc Natl Acad Sci U S A. 2011;108(13):5209-14.

59. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328(5979):710-22.

60. Hoffmann DL, Standish CD, García-Diez M, Pettitt PB, Milton JA, Zilhão J, et al. U-Th dating of carbonate crusts reveals Neandertal origin of Iberian cave art. Science. 2018;359(6378):912.

61. Gokhman D, Mishol N, de Manuel M, de Juan D, Shuqrun J, Meshorer E, et al. Reconstructing Denisovan Anatomy Using DNA Methylation Maps. Cell. 2019;179(1):180-92.e10.

62. DeGiorgio M, Jakobsson M, Rosenberg NA. Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. Proc Natl Acad Sci U S A. 2009;106(38):16057-62.

63. Jobling MA. Human evolutionary genetics. New York and London: Garland Science; 2014.

64. Klein RG. The archeology of modern human origins. Evolutionary Anthropology: Issues, News, and Reviews. 1992;1(1):5-14.

65. McBrearty S, Brooks AS. The revolution that wasn't: a new interpretation of the origin of modern human behavior. J Hum Evol. 2000;39(5):453-563.

66. Beaudet A. The Emergence of Language in the Hominin Lineage: Perspectives from Fossil Endocasts. Frontiers in Human Neuroscience. 2017;11:427.

67. Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain.

68. Wonders CP, Anderson SA. The origin and specification of cortical interneurons. Nat Rev Neurosci. 2006;7(9):687-96.

69. Douglas RJ, Martin KA. Neuronal circuits of the neocortex. Annu Rev Neurosci. 2004;27:419-51.

70. Lopez-Bendito G, Molnar Z. Thalamocortical development: how are we going to get there? Nat Rev Neurosci. 2003;4(4):276-89.

71. Kaas JH. The Evolution of Brains from Early Mammals to Humans. Wiley Interdiscip Rev Cogn Sci. 2013;4(1):33-45.

72. Götz M. Cerebral Cortex Development. eLS. 2001.

73. Ackerman S, National Institute of Mental H. Discovering the brain. 1997.

74. Hébert JM, Fishell G. The genetics of early telencephalon patterning: some assembly required. Nat Rev Neurosci. 2008;9(9):678-85.

75. Bystron I, Blakemore C, Rakic P. Development of the human cerebral cortex: Boulder Committee revisited. Nature reviews Neuroscience. 2008;9(2):110-22.

76. Rakic P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. Trends in neurosciences. 1995;18(9):383-8.

77. Florio M, Huttner WB. Neural progenitors, neurogenesis and the evolution of the neocortex. Development (Cambridge, England). 2014;141(11):2182-94.

78. Embryonic vertebrate central nervous system: revised terminology. The Boulder Committee. Anat Rec. 1970;166(2):257-61.

79. Zhao S, Frotscher M. Go or Stop? Divergent Roles of Reelin in Radial Neuronal Migration. The Neuroscientist. 2010;16(4):421-34.

80. Kilb W, Frotscher M. Cajal-Retzius cells: organizers of cortical development. e-Neuroforum. 2016;7(4):82-8.

81. Jiang X, Nardelli J. Cellular and molecular introduction to brain development. Neurobiol Dis. 2016;92(Pt A):3-17.

82. Nicholls JG, Kuffler SW. From neuron to brain2018.

83. Nelson CA. Threats to optimal development : integrating biological, psychological, and social risk factors2018.

84. Ruszczycki B, Szepesi Z, Wilczynski GM, Bijata M, Kalita K, Kaczmarek L, et al. Sampling issues in quantitative analysis of dendritic spines morphology. BMC bioinformatics. 2012;13:213-.

85. Calabrese B, Wilson MS, Halpain S. Development and regulation of dendritic spine synapses. Physiology (Bethesda, Md). 2006;21:38-47.

86. Nimchinsky EA, Sabatini BL, Svoboda K. Structure and function of dendritic spines. Annual review of physiology. 2002;64:313-53.

87. Silbereis JC, Pochareddy S, Zhu Y, Li M, Sestan N. The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. Neuron. 2016;89(2):248-68.

88. Tau GZ, Peterson BS. Normal Development of Brain Circuits. Neuropsychopharmacology. 2010;35(1):147-68.

89. Huttenlocher PR, Dabholkar AS. Regional differences in synaptogenesis in human cerebral cortex. J Comp Neurol. 1997;387(2):167-78.

90. Wilson ES, Newell-Litwa K. Stem cell models of human synapse development and degeneration. Molecular Biology of the Cell. 2018;29(24):2913-21.

91. Pattabiraman K, Muchnik SK, Sestan N. The evolution of the human brain and disease susceptibility. Current opinion in genetics & development. 2020;65:91-7.

92. Won H, Mah W, Kim E. Autism spectrum disorder causes, mechanisms, and treatments: focus on neuronal synapses. Frontiers in Molecular Neuroscience. 2013;6:19.

93. Gao R, Penzes P. Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. Curr Mol Med. 2015;15(2):146-67.

94. Habela CW, Song H, Ming G-l. Modeling synaptogenesis in schizophrenia and autism using human iPSC derived neurons. Molecular and Cellular Neuroscience. 2016;73:52-62.

95. Saha S, Chant D, Welham J, McGrath J. A Systematic Review of the Prevalence of Schizophrenia. PLOS Medicine. 2005;2(5):e141.

96. Brown S. Excess mortality of schizophrenia. A meta-analysis. Br J Psychiatry. 1997;171:502-8.

97. Alexander-Bloch AF, Vértes PE, Stidd R, Lalonde F, Clasen L, Rapoport J, et al. The Anatomical Distance of Functional Connections Predicts Brain Network Topology in Health and Schizophrenia. Cerebral Cortex. 2012;23(1):127-38.

98. Kolluri N, Sun Z, Sampson AR, Lewis DA. Lamina-Specific Reductions in Dendritic Spine Density in the Prefrontal Cortex of Subjects With Schizophrenia. American Journal of Psychiatry. 2005;162(6):1200-2.

99. Sweet RA, Henteleff RA, Zhang W, Sampson AR, Lewis DA. Reduced Dendritic Spine Density in Auditory Cortex of Subjects with Schizophrenia. Neuropsychopharmacology. 2009;34(2):374-89.

100. Garey LJ, Ong WY, Patel TS, Kanani M, Davis A, Mortimer AM, et al. Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia. Journal of Neurology, Neurosurgery & Psychiatry. 1998;65(4):446.

101. Glantz LA, Lewis DA. Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. Arch Gen Psychiatry. 2000;57(1):65-73.

102. Broadbelt K, Byne W, Jones LB. Evidence for a decrease in basilar dendrites of pyramidal cells in schizophrenic medial prefrontal cortex. Schizophrenia Research. 2002;58(1):75-81.

103. Hu W, MacDonald ML, Elswick DE, Sweet RA. The glutamate hypothesis of schizophrenia: evidence from human brain tissue studies. Annals of the New York Academy of Sciences. 2015;1338(1):38-57.

104. Mistry M, Gillis J, Pavlidis P. Meta-analysis of gene coexpression networks in the post-mortem prefrontal cortex of patients with schizophrenia and unaffected controls. BMC Neuroscience. 2013;14(1):105.

105. Lai M-C, Lombardo MV, Baron-Cohen S. Autism. The Lancet. 2014;383(9920):896-910.

106. Huguet G, Ey E, Bourgeron T. The Genetic Landscapes of Autism Spectrum Disorders. Annual Review of Genomics and Human Genetics. 2013;14(1):191-213.

107. Hutsler JJ, Zhang H. Increased dendritic spine densities on cortical projection neurons in autism spectrum disorders. Brain Res. 2010;1309:83-94.

108. Irwin SA, Patel B, Idupulapati M, Harris JB, Crisostomo RA, Larsen BP, et al. Abnormal dendritic spine characteristics in the temporal and visual cortices of

patients with fragile-X syndrome: a quantitative examination. Am J Med Genet. 2001;98(2):161-7.

109. Gilbert J, Man H-Y. Fundamental Elements in Autism: From Neurogenesis and Neurite Growth to Synaptic Plasticity. Frontiers in Cellular Neuroscience. 2017;11:359.

110. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180(3):568-84.e23.

111. Kulkarni VA, Firestein BL. The dendritic tree and brain disorders. Molecular and Cellular Neuroscience. 2012;50(1):10-20.

112. Brennand KJ, Marchetto MC, Benvenisty N, Brustle O, Ebert A, Izpisua Belmonte JC, et al. Creating Patient-Specific Neural Cells for the In Vitro Study of Brain Disorders. Stem Cell Reports. 2015;5(6):933-45.

113. Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, et al. Cerebral organoids model human brain development and microcephaly. Nature. 2013;501(7467):373-9.

114. Finlay BL, Darlington RB. Linked regularities in the development and evolution of mammalian brains. Science (New York, NY). 1995;268(5217):1578-84.

115. Sherwood CC, Subiaul F, Zawidzki TW. A natural history of the human mind: tracing evolutionary changes in brain and cognition. Journal of Anatomy. 2008;212(4):426-54.

116. Sherwood CC, Stimpson CD, Raghanti MA, Wildman DE, Uddin M, Grossman LI, et al. Evolution of increased glia–neuron ratios in the human frontal cortex. Proceedings of the National Academy of Sciences. 2006;103(37):13606.

117. Semendeferi K, Teffer K, Buxhoeveden DP, Park MS, Bludau S, Amunts K, et al. Spatial organization of neurons in the frontal pole sets humans apart from great apes. Cerebral cortex (New York, NY : 1991). 2011;21(7):1485-97.

118. Liu Y, Konopka G. An integrative understanding of comparative cognition: lessons from human brain evolution. Integrative and Comparative Biology. 2020.

119. Mountcastle V. The evolution of ideas concerning the function of the neocortex. Cerebral cortex (New York, NY : 1991). 1995;5(4):289-95.

120. Goldman-Rakic PS. Development of cortical circuitry and cognitive function. Child development. 1987;58(3):601-22.

121. LaMonica BE, Lui JH, Wang X, Kriegstein AR. OSVZ progenitors in the human cortex: an updated perspective on neurodevelopmental disease. Current opinion in neurobiology. 2012;22(5):747-53.

122. Kasai H, Fukuda M, Watanabe S, Hayashi-Takagi A, Noguchi J. Structural dynamics of dendritic spines in memory and cognition. Trends Neurosci. 2010;33(3):121-9.

123. Elston GN, Benavides-Piccione R, DeFelipe J. The pyramidal cell in cognition: a comparative study in human and monkey. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2001;21(17):Rc163.

124. Sherwood CC, Miller SB, Karl M, Stimpson CD, Phillips KA, Jacobs B, et al. Invariant Synapse Density and Neuronal Connectivity Scaling in Primate Neocortical Evolution. Cerebral Cortex. 2020.

125. Anderson JM, Gilmore R, Roper S, Crosson B, Bauer RM, Nadeau S, et al. Conduction Aphasia and the Arcuate Fasciculus: A Reexamination of the Wernicke–Geschwind Model. Brain and Language. 1999;70(1):1-12.

126. Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, et al. The evolution of the arcuate fasciculus revealed with comparative DTI. Nature Neuroscience. 2008;11(4):426-8.

127. Mantini D, Corbetta M, Romani GL, Orban GA, Vanduffel W. Evolutionarily Novel Functional Networks in the Human Brain? The Journal of Neuroscience. 2013;33(8):3259.

128. Hecht EE, Gutman DA, Bradley BA, Preuss TM, Stout D. Virtual dissection and comparative connectivity of the superior longitudinal fasciculus in chimpanzees and humans. NeuroImage. 2015;108:124-37.

129. Allman JM, Tetreault NA, Hakeem AY, Manaye KF, Semendeferi K, Erwin JM, et al. The von Economo neurons in frontoinsular and anterior cingulate cortex in great apes and humans. Brain Structure and Function. 2010;214(5):495-517.

130. Oberheim NA, Goldman SA, Nedergaard M. Heterogeneity of Astrocytic Form and Function. In: Milner R, editor. Astrocytes: Methods and Protocols. Totowa, NJ: Humana Press; 2012. p. 23-45.

131. Blanco-Suárez E, Caldwell ALM, Allen NJ. Role of astrocyte–synapse interactions in CNS disorders. The Journal of Physiology. 2017;595(6):1903-16.

132. Zhang Y, Barres BA. A smarter mouse with human astrocytes. Bioessays. 2013;35(10):876-80.

133. Miller DJ, Duka T, Stimpson CD, Schapiro SJ, Baze WB, McArthur MJ, et al. Prolonged myelination in human neocortical evolution. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(41):16480-5.

134. Petanjek Z, Judas M, Simic G, Rasin MR, Uylings HB, Rakic P, et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(32):13281-6.

135. Elston GN, Oga T, Fujita I. Spinogenesis and pruning scales across functional hierarchies. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009;29(10):3271-5.

136. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 2001;413(6855):519-23.

137. Hurst JA, Baraitser M, Auger E, Graham F, Norell S. An extended family with a dominantly inherited speech disorder. Dev Med Child Neurol. 1990;32(4):352-5.

138. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, et al. Molecular evolution of FOXP2, a gene involved in speech and language. Nature. 2002;418(6900):869-72.

139. Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, et al. The derived FOXP2 variant of modern humans was shared with Neandertals. Curr Biol. 2007;17(21):1908-12.

140. Atkinson EG, Audesse AJ, Palacios JA, Bobo DM, Webb AE, Ramachandran S, et al. No Evidence for Recent Selection at FOXP2 among Diverse Human Populations. Cell. 2018;174(6):1424-35.e15.

141. Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, et al. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell. 2009;137(5):961-71.

142. Wohlgemuth S, Adam I, Scharff C. FoxP2 in songbirds. Current opinion in neurobiology. 2014;28:86-93.

143. Teramitsu I, Poopatanapong A, Torrisi S, White SA. Striatal FoxP2 is actively regulated during songbird sensorimotor learning. PLoS One. 2010;5(1):e8548.

144. Teramitsu I, White SA. FoxP2 regulation during undirected singing in adult songbirds. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2006;26(28):7390-4.

145. Enard W. FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. Current opinion in neurobiology. 2011;21(3):415-24.

146. Schulz SB, Haesler S, Scharff C, Rochefort C. Knockdown of FoxP2 alters spine density in Area X of the zebra finch. Genes, brain, and behavior. 2010;9(7):732-40.

147. Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, et al. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS genetics. 2011;7(7):e1002145.

148. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. Nature reviews Genetics. 2009;10(10):725-32.

149. Conrad B, Antonarakis SE. Gene duplication: a drive for phenotypic diversity and cause of human disease. Annual review of genomics and human genetics. 2007;8:17-35.

150. Dennis MY, Nuttle X, Sudmant PH, Antonacci F, Graves TA, Nefedov M, et al. Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. Cell. 2012;149(4):912-22.

151. Charrier C, Joshi K, Coutinho-Budd J, Kim JE, Lambert N, de Marchena J, et al. Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. Cell. 2012;149(4):923-35.

152. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. Genome biology. 2004;5(10):R74.

153. Calarco JA, Xing Y, Cáceres M, Calarco JP, Xiao X, Pan Q, et al. Global analysis of alternative splicing differences between humans and chimpanzees. Genes & Development. 2007;21(22):2963-75.

154. Wang W, Nahta R, Huper G, Marks JR. TAFII70 isoform-specific growth suppression correlates with its ability to complex with the GADD45a protein. Molecular cancer research : MCR. 2004;2(8):442-52.

155. Wang S, Dibenedetto AJ, Pittman RN. Genes induced in programmed cell death of neuronal PC12 cells and developing sympathetic neurons in vivo. Developmental biology. 1997;188(2):322-36.

156. Reilly SK, Yin J, Ayoub AE, Emera D, Leng J, Cotney J, et al. Evolutionary Changes in Promoter and Enhancer Activity During Human Corticogenesis. Science (New York, NY). 2015;347(6226):1155-9.

157. Shulha HP, Crisci JL, Reshetov D, Tushir JS, Cheung I, Bharadwaj R, et al. Human-Specific Histone Methylation Signatures at Transcription Start Sites in Prefrontal Neurons. PLoS biology. 2012;10(11).

158. Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H, Prado-Martinez J, et al. The interplay between DNA methylation and sequence divergence in recent human evolution. Nucleic acids research. 2015;43(17):8204-14.

159. Mendizabal I, Shi L, Keller TE, Konopka G, Preuss TM, Hsieh TF, et al. Comparative Methylome Analyses Identify Epigenetic Regulatory Loci of Human Brain Evolution. Molecular biology and evolution. 2016;33(11):2947-59.

160. Zeng J, Konopka G, Hunt B G, Preuss T M, Geschwind D, Yi S V. Divergent Whole-Genome Methylation Maps of Human and Chimpanzee Brains Reveal Epigenetic Basis of Human Regulatory Evolution. American journal of human genetics. 2012;91(3):455-65.

161. Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, et al. Dynamics of DNA methylation in recent human and great ape evolution. PLoS genetics. 2013;9(9):e1003763.

162. Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. PLoS genetics. 2011;7(2):e1001316.

163. Berezikov E. Evolution of microRNA diversity and regulation in animals. Nature reviews Genetics. 2011;12(12):846-60.

164. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 2006;443(7108):167-72.

165. Hubisz MJ, Pollard KS. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. Current Opinion in Genetics & Development. 2014;29:15-21.

166. Burbano HA, Green RE, Maricic T, Lalueza-Fox C, de la Rasilla M, Rosas A, et al. Analysis of human accelerated DNA regions using archaic hominin genomes. PLoS One. 2012;7(3):e32877.

167. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013;47:97-120.

168. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. Science. 2006;312(5780):1614-20.

169. Kostka D, Hubisz MJ, Siepel A, Pollard KS. The Role of GC-Biased Gene Conversion in Shaping the Fastest Evolving Regions of the Human Genome. Molecular Biology and Evolution. 2011;29(3):1047-57.

170. Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. Many human accelerated regions are developmental enhancers. Philos Trans R Soc Lond B Biol Sci. 2013;368(1632):20130025.

171. Levchenko A, Kanapin A, Samsonova A, Gainetdinov RR. Human Accelerated Regions and Other Human-Specific Sequence Variations in the Context of Evolution and Their Relevance for Brain Development. Genome Biol Evol. 2018;10(1):166-88.

172. Boyd JL, Skove SL, Rouanet JP, Pilaz LJ, Bepler T, Gordan R, et al. Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. Curr Biol. 2015;25(6):772-9.

173. Shi Y, Kirwan P, Smith J, Robinson HP, Livesey FJ. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nature neuroscience. 2012;15(3):477-86, s1.

174. Xu K, Schadt EE, Pollard KS, Roussos P, Dudley JT. Genomic and Network Patterns of Schizophrenia Genetic Variation in Human Evolutionary Accelerated Regions. Molecular Biology and Evolution. 2015;32(5):1148-60.

175. Won H, Huang J, Opland CK, Hartl CL, Geschwind DH. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. Nat Commun. 2019;10(1):2396.

176. Duret L, Galtier N. Comment on "Human-specific gain of function in a developmental enhancer". Science. 2009;323(5915):714; author reply

177. Sumiyama K, Saitou N. Loss-of-Function Mutation in a Repressor Module of Human-Specifically Activated Enhancer HACNS1. Molecular Biology and Evolution. 2011;28(11):3005-7.

178. Lewis DA. The Human Brain Revisited: Opportunities and Challenges in Postmortem Studies of Psychiatric Disorders. Neuropsychopharmacology. 2002;26(2):143-54.

179. Stan AD, Ghose S, Gao XM, Roberts RC, Lewis-Amezcua K, Hatanpaa KJ, et al. Human postmortem tissue: what quality markers matter? Brain Res. 2006;1123(1):1-11.

180. Suzuki IK, Vanderhaeghen P. Is this a brain which I see before me? Modeling human neural development with pluripotent stem cells. Development. 2015;142(18):3138-50.

181. Shi Y, Kirwan P, Livesey FJ. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nature protocols. 2012;7(10):1836-46.

182. Siegenthaler JA, Ashique AM, Zarbalis K, Patterson KP, Hecht JH, Kane MA, et al. Retinoic Acid from the Meninges Regulates Cortical Neuron Generation. Cell. 2009;139(3):597-609.

183. Zhou J, Su P, Li D, Tsang S, Duan E, Wang F. High-efficiency induction of neural conversion in human ESCs and human induced pluripotent stem cells with a single chemical inhibitor of transforming growth factor beta superfamily receptors. Stem Cells. 2010;28(10):1741-50.

184. Borooah S, Phillips MJ, Bilican B, Wright AF, Wilmut I, Chandran S, et al. Using human induced pluripotent stem cells to treat retinal disease. Progress in Retinal and Eye Research. 2013;37:163-81.

185. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009;27(3):275-80.

186. Kirwan P, Turner-Bridger B, Peter M, Momoh A, Arambepola D, Robinson HP, et al. Development and function of human cerebral cortex neural networks from pluripotent stem cells in vitro. Development. 2015;142(18):3178-87.

187. Strano A, Tuck E, Stubbs VE, Livesey FJ. Variable Outcomes in Neural Differentiation of Human PSCs Arise from Intrinsic Differences in Developmental Signaling Pathways. Cell Rep. 2020;31(10):107732.

188. Workman AD, Charvet CJ, Clancy B, Darlington RB, Finlay BL. Modeling Transformations of Neurodevelopmental Sequences across Mammalian Species. The Journal of Neuroscience. 2013;33(17):7368.

189. Rodenas-Cuadrado P, Pietrafusa N, Francavilla T, La Neve A, Striano P, Vernes SC. Characterisation of CASPR2 deficiency disorder--a syndrome involving autism, epilepsy and language impairment. BMC Med Genet. 2016;17:8.

190. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.

191. Poliak S, Gollan L, Martinez R, Custer A, Einheber S, Salzer JL, et al. Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K+ channels. Neuron. 1999;24(4):1037-47.

192. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, et al. Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. N Engl J Med. 2006;354(13):1370-7.

193. Poliak S, Salomon D, Elhanany H, Sabanay H, Kiernan B, Pevny L, et al. Juxtaparanodal clustering of Shaker-like K+ channels in myelinated axons depends on Caspr2 and TAG-1. J Cell Biol. 2003;162(6):1149-60.

194. Scott R, Sanchez-Aguilera A, van Elst K, Lim L, Dehorter N, Bae SE, et al. Loss of Cntnap2 Causes Axonal Excitability Deficits, Developmental Delay in Cortical Myelination, and Abnormal Stereotyped Motor Behavior. Cereb Cortex. 2019;29(2):586-97.

195. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, et al. A functional genetic link between distinct developmental language disorders. N Engl J Med. 2008;359(22):2337-45.

196. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, et al. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet. 2008;82(1):165-73.

197. Abrahams BS, Tentler D, Perederiy JV, Oldham MC, Coppola G, Geschwind DH. Genome-wide analyses of human perisylvian cerebral cortical patterning. Proc Natl Acad Sci U S A. 2007;104(45):17849-54.

198. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. Conserved cell types with divergent features in human versus mouse cortex. Nature. 2019;573(7772):61-8.

199. Vogt D, Cho KKA, Shelton SM, Paul A, Huang ZJ, Sohal VS, et al. Mouse Cntnap2 and Human CNTNAP2 ASD Alleles Cell Autonomously Regulate PV+ Cortical Interneurons. Cereb Cortex. 2018;28(11):3868-79.

200. Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet. 2008;82(1):150-9.

201. Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D, et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron. 2009;62(4):494-509.

202. Gao R, Piguel NH, Melendez-Zaidi AE, Martin-de-Saavedra MD, Yoon S, Forrest MP, et al. CNTNAP2 stabilizes interneuron dendritic arbors through CASK. Mol Psychiatry. 2018;23(9):1832-50.

203. Fernandes D, Santos SD, Coutinho E, Whitt JL, Beltrao N, Rondao T, et al. Disrupted AMPA Receptor Function upon Genetic- or Antibody-Mediated Loss of Autism-Associated CASPR2. Cereb Cortex. 2019;29(12):4919-31.

204. Belloso JM, Bache I, Guitart M, Caballin MR, Halgren C, Kirchhoff M, et al. Disruption of the CNTNAP2 gene in a t(7;15) translocation family without symptoms of Gilles de la Tourette syndrome. Eur J Hum Genet. 2007;15(6):711-3.

205. Watson CM, Crinnion LA, Tzika A, Mills A, Coates A, Pendlebury M, et al. Diagnostic whole genome sequencing and split-read mapping for nucleotide resolution breakpoint identification in CNTNAP2 deficiency syndrome. Am J Med Genet A. 2014;164a(10):2649-55.

206. Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, et al. Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron. 2015;88(3):499-513.

207. Smogavec M, Cleall A, Hoyer J, Lederer D, Nassogne MC, Palmer EE, et al. Eight further individuals with intellectual disability and epilepsy carrying bi-allelic CNTNAP2 aberrations allow delineation of the mutational and phenotypic spectrum. J Med Genet. 2016;53(12):820-7.

208. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, et al. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. Am J Hum Genet. 2008;82(1):160-4.

209. Rossi E, Verri AP, Patricelli MG, Destefani V, Ricca I, Vetro A, et al. A 12Mb deletion at 7q33-q35 associated with autism spectrum disorders and primary amenorrhea. Eur J Med Genet. 2008;51(6):631-8.

210. Li X, Hu Z, He Y, Xiong Z, Long Z, Peng Y, et al. Association analysis of CNTNAP2 polymorphisms with autism in the Chinese Han population. Psychiatr Genet. 2010;20(3):113-7.

211. Steer CD, Golding J, Bolton PF. Traits contributing to the autistic spectrum. PLoS One. 2010;5(9):e12633.

212. Nord AS, Roeb W, Dickel DE, Walsh T, Kusenda M, O'Connor KL, et al. Reduced transcript expression of genes affected by inherited and de novo CNVs in autism. Eur J Hum Genet. 2011;19(6):727-31.

213. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011;43(6):585-9.

214. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum Mol Genet. 2012;21(21):4781-92.

215. Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, Sato D, et al. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. G3 (Bethesda). 2012;2(12):1665-85.

216. Sampath S, Bhat S, Gupta S, O'Connor A, West AB, Arking DE, et al. Defining the contribution of CNTNAP2 to autism susceptibility. PLoS One. 2013;8(10):e77906.

217. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am J Hum Genet. 2013;92(2):221-37.

218. Koshimizu E, Miyatake S, Okamoto N, Nakashima M, Tsurusaki Y, Miyake N, et al. Performance comparison of bench-top next generation sequencers using microdroplet PCR-based enrichment for targeted sequencing in patients with autism spectrum disorder. PLoS One. 2013;8(9):e74167.

219. Chiocchetti AG, Kopp M, Waltes R, Haslinger D, Duketis E, Jarczok TA, et al. Variants of the CNTNAP2 5' promoter as risk factors for autism spectrum disorders: a genetic and functional approach. Mol Psychiatry. 2015;20(7):839-49.

220. Eriksson MA, Lieden A, Westerlund J, Bremer A, Wincent J, Sahlin E, et al. Rare copy number variants are common in young children with autism spectrum disorder. Acta Paediatr. 2015;104(6):610-8.

221. Nascimento PP, Bossolani-Martins AL, Rosan DB, Mattos LC, Brandao-Mattos C, Fett-Conte AC. Single nucleotide polymorphisms in the CNTNAP2 gene in Brazilian patients with autistic spectrum disorder. Genet Mol Res. 2016;15(1).

222. Zhou WZ, Zhang J, Li Z, Lin X, Li J, Wang S, et al. Targeted resequencing of 358 candidate genes for autism spectrum disorder in a Chinese cohort reveals diagnostic potential and genotype-phenotype correlations. Hum Mutat. 2019;40(6):801-15.

223. Aspromonte MC, Bellini M, Gasparini A, Carraro M, Bettella E, Polli R, et al. Characterization of intellectual disability and autism comorbidity through gene panel sequencing. Hum Mutat. 2019;40(9):1346-63.

224. Egger G, Roetzer KM, Noor A, Lionel AC, Mahmood H, Schwarzbraun T, et al. Identification of risk genes for autism spectrum disorder through copy number variation analysis in Austrian families. Neurogenetics. 2014;15(2):117-27.

225. Petrin AL, Giacheti CM, Maximino LP, Abramides DV, Zanchetta S, Rossi NF, et al. Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case. Am J Med Genet A. 2010;152a(12):3164-72.

226. Newbury DF, Paracchini S, Scerri TS, Winchester L, Addis L, Richardson AJ, et al. Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. Behav Genet. 2011;41(1):90-104.

227. Stein MB, Yang BZ, Chavira DA, Hitchcock CA, Sung SC, Shipon-Blum E, et al. A common genetic variant in the neurexin superfamily member CNTNAP2 is associated with increased risk for selective mutism and social anxiety-related traits. Biol Psychiatry. 2011;69(9):825-31.

228. Al-Murrani A, Ashton F, Aftimos S, George AM, Love DR. Amino-Terminal Microdeletion within the CNTNAP2 Gene Associated with Variable Expressivity of Speech Delay. Case Rep Genet. 2012;2012:172408.

229. Laffin JJ, Raca G, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD. Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. Genet Med. 2012;14(11):928-36.

230. Centanni TM, Sanmann JN, Green JR, Iuzzini-Seigel J, Bartlett C, Sanger WG, et al. The role of candidate-gene CNTNAP2 in childhood apraxia of speech and specific language impairment. Am J Med Genet B Neuropsychiatr Genet. 2015;168(7):536-43.

231. Chen XS, Reader RH, Hoischen A, Veltman JA, Simpson NH, Francks C, et al. Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. Sci Rep. 2017;7:46105.

232. Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet. 2010;6(5):e1000962.

233. Lesca G, Rudolf G, Labalme A, Hirsch E, Arzimanoglou A, Genton P, et al. Epileptic encephalopathies of the Landau-Kleffner and continuous spike and waves during slow-wave sleep types: genomic dissection makes the link with autism. Epilepsia. 2012;53(9):1526-38.

234. Pippucci T, Licchetta L, Baldassari S, Palombo F, Menghi V, D'Aurizio R, et al. Epilepsy with auditory features: A heterogeneous clinico-molecular disease. Neurol Genet. 2015;1(1):e5.

235. Fisher SE, Lai CS, Monaco AP. Deciphering the genetic basis of speech and language disorders. Annu Rev Neurosci. 2003;26:57-80.

236. Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, et al. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry. 2008;13(3):261-6.

237. Clemm von Hohenberg C, Wigand MC, Kubicki M, Leicht G, Giegling I, Karch S, et al. CNTNAP2 polymorphisms and structural brain connectivity: a diffusion-tensor imaging study. J Psychiatr Res. 2013;47(10):1349-56.

238. Lee IS, Carvalho CM, Douvaras P, Ho SM, Hartley BJ, Zuccherato LW, et al. Characterization of molecular and cellular phenotypes associated with a heterozygous CNTNAP2 deletion using patient-derived hiPSC neural cells. NPJ Schizophr. 2015;1.

239. Flaherty E, Deranieh RM, Artimovich E, Lee IS, Siegel AJ, Levy DL, et al. Patient-derived hiPSC neurons with heterozygous CNTNAP2 deletions display altered neuronal gene expression and network activity. NPJ Schizophr. 2017;3(1):35.

240. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry. 2010;15(6):637-46.

241. Verkerk AJ, Mathews CA, Joosse M, Eussen BH, Heutink P, Oostra BA. CNTNAP2 is disrupted in a family with Gilles de la Tourette syndrome and obsessive compulsive disorder. Genomics. 2003;82(1):1-9.

242. Peter B, Raskind WH, Matsushita M, Lisowski M, Vu T, Berninger VW, et al. Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. J Neurodev Disord. 2011;3(1):39-49.

243. Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB. Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity. J Hum Genet. 2013;58(8):539-47.

244. Jackman C, Horn ND, Molleston JP, Sokol DK. Gene associated with seizures, autism, and hepatomegaly in an Amish girl. Pediatr Neurol. 2009;40(4):310-3.

245. Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell. 2011;147(1):235-46.

246. Sweatt JD. Pitt–Hopkins Syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. Exp Mol Med. 2013;45(5):e21-.

247. Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, et al. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. Am J Hum Genet. 2009;85(5):655-66.

248. Riccardi F, Urquhart J, McCullagh G, Lawrence P, Douzgou S. A patient with a novel CNTNAP2 homozygous variant: further delineation of the CASPR2 deficiency syndrome and review of the literature. Clin Dysmorphol. 2019;28(2):66-70.

249. Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, Engels H, et al. Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. BMC Med Genet. 2011;12:106.

250. Parrini E, Marini C, Mei D, Galuppi A, Cellini E, Pucatti D, et al. Diagnostic Targeted Resequencing in 349 Patients with Drug-Resistant Pediatric Epilepsies Identifies Causative Mutations in 30 Different Genes. Hum Mutat. 2017;38(2):216-25.

251. Tan GC, Doke TF, Ashburner J, Wood NW, Frackowiak RS. Normal variation in fronto-occipital circuitry and cerebellar structure with an autism-associated polymorphism of CNTNAP2. Neuroimage. 2010;53(3):1030-42.

252. Toma C, Pierce KD, Shaw AD, Heath A, Mitchell PB, Schofield PR, et al. Comprehensive cross-disorder analyses of CNTNAP2 suggest it is unlikely to be a primary risk gene for psychiatric disorders. PLoS Genet. 2018;14(12):e1007535.

253. Huckins LM, Dobbyn A, Ruderfer DM, Hoffman G, Wang W, Pardiñas AF, et al. Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. Nature Genetics. 2019;51(4):659-74.

254. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA, et al. Genetic identification of brain cell types underlying schizophrenia. Nature Genetics. 2018;50(6):825-33.

255. Whalley HC, O'Connell G, Sussmann JE, Peel A, Stanfield AC, Hayiou-Thomas ME, et al. Genetic variation in CNTNAP2 alters brain function during linguistic processing in healthy individuals. Am J Med Genet B Neuropsychiatr Genet. 2011;156b(8):941-8.

256. Whitehouse AJ, Bishop DV, Ang QW, Pennell CE, Fisher SE. CNTNAP2 variants affect early language development in the general population. Genes Brain Behav. 2011;10(4):451-6.

257. Scott-Van Zeeland AA, Abrahams BS, Alvarez-Retuerto AI, Sonnenblick LI, Rudie JD, Ghahremani D, et al. Altered functional connectivity in frontal lobe circuits is associated with variation in the autism risk gene CNTNAP2. Sci Transl Med. 2010;2(56):56ra80.

258. Dennis EL, Jahanshad N, Rudie JD, Brown JA, Johnson K, McMahon KL, et al. Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2. Brain Connect. 2011;1(6):447-59.

259. Ji W, Li T, Pan Y, Tao H, Ju K, Wen Z, et al. CNTNAP2 is significantly associated with schizophrenia and major depression in the Han Chinese population. Psychiatry Res. 2013;207(3):225-8.

260. Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, et al. Single-cell genomics identifies cell type-specific molecular changes in autism. Science. 2019;364(6441):685-9.

261. Canali G, Garcia M, Hivert B, Pinatel D, Goullancourt A, Oguievetskaia K, et al. Genetic variants in autism-related CNTNAP2 impair axonal growth of cortical neurons. Hum Mol Genet. 2018;27(11):1941-54.

262. Kim JW, Park K, Kang RJ, Gonzales ELT, Kim DG, Oh HA, et al. Pharmacological modulation of AMPA receptor rescues social impairments in animal models of autism. Neuropsychopharmacology. 2019;44(2):314-23.

263. Lazaro MT, Taxidis J, Shuman T, Bachmutsky I, Ikrar T, Santos R, et al. Reduced Prefrontal Synaptic Connectivity and Disturbed Oscillatory Population Dynamics in the CNTNAP2 Model of Autism. Cell Rep. 2019;27(9):2567-78.e6.

264. Liska A, Bertero A, Gomolka R, Sabbioni M, Galbusera A, Barsotti N, et al. Homozygous Loss of Autism-Risk Gene CNTNAP2 Results in Reduced Local and Long-Range Prefrontal Functional Connectivity. Cereb Cortex. 2018;28(4):1141-53.

265. Zerbi V, Ielacqua GD, Markicevic M, Haberl MG, Ellisman MH, A AB, et al. Dysfunctional Autism Risk Genes Cause Circuit-Specific Connectivity Deficits With Distinct Developmental Trajectories. Cereb Cortex. 2018;28(7):2495-506.

266. Herbert MR, Ziegler DA, Deutsch CK, O'Brien LM, Kennedy DN, Filipek PA, et al. Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. Brain. 2005;128(Pt 1):213-26.

267. Courchesne E, Pierce K. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. Curr Opin Neurobiol. 2005;15(2):225-30.

268. Gordon A, Salomon D, Barak N, Pen Y, Tsoory M, Kimchi T, et al. Expression of Cntnap2 (Caspr2) in multiple levels of sensory systems. Mol Cell Neurosci. 2016;70:42-53.

269. Hoffman EJ, Turner KJ, Fernandez JM, Cifuentes D, Ghosh M, Ijaz S, et al. Estrogens Suppress a Behavioral Phenotype in Zebrafish Mutants of the Autism Risk Gene, CNTNAP2. Neuron. 2016;89(4):725-33.

270. Thomas AM, Schwartz MD, Saxe MD, Kilduff TS. Cntnap2 Knockout Rats and Mice Exhibit Epileptiform Activity and Abnormal Sleep-Wake Physiology. Sleep. 2017;40(1).

271. Selimbeyoglu A, Kim CK, Inoue M, Lee SY, Hong ASO, Kauvar I, et al. Modulation of prefrontal cortex excitation/inhibition balance rescues social behavior in CNTNAP2-deficient mice. Sci Transl Med. 2017;9(401).

272. Hali S, Kim J, Kwak TH, Lee H, Shin CY, Han DW. Modelling monogenic autism spectrum disorder using mouse cortical organoids. Biochem Biophys Res Commun. 2020;521(1):164-71.

273. Brunner D, Kabitzke P, He D, Cox K, Thiede L, Hanania T, et al. Comprehensive Analysis of the 16p11.2 Deletion and Null Cntnap2 Mouse Models of Autism Spectrum Disorder. PLoS One. 2015;10(8):e0134572.

274. Chorev M, Carmel L. The function of introns. Front Genet. 2012;3:55.

275. Gittelman RM, Hun E, Ay F, Madeoy J, Pennacchio L, Noble WS, et al. Comprehensive identification and analysis of human accelerated regulatory DNA. Genome Res. 2015;25(9):1245-55.

276. Capra JA, Hubisz MJ, Kostka D, Pollard KS, Siepel A. A Model-Based Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes. PLOS Genetics. 2013;9(8):e1003684.

277. Dorschner MO, Hawrylycz M, Humbert R, Wallace JC, Shafer A, Kawamoto J, et al. High-throughput localization of functional elements by quantitative chromatin profiling. Nat Methods. 2004;1(3):219-25.

278. Adam I, Mendoza E, Kobalz U, Wohlgemuth S, Scharff C. CNTNAP2 is a direct FoxP2 target in vitro and in vivo in zebra finches: complex regulation by age and activity. Genes Brain Behav. 2017;16(6):635-42.

279. Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature. 2012;482(7384):216-20.

280. Bruntraeger M, Byrne M, Long K, Bassett AR. Editing the Genome of Human Induced Pluripotent Stem Cells Using CRISPR/Cas9 Ribonucleoprotein Complexes. Methods Mol Biol. 2019;1961:153-83.

281. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biology. 2018;19(1):224.

282. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33(5):495-502.

283. Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, et al. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 2016;538(7624):238-42.

284. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5.

285. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886-d94.

286. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

287. Polioudakis D, de la Torre-Ubieta L, Langerman J, Elkins AG, Shi X, Stein JL, et al. A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron. 2019;103(5):785-801.e8.

288. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012;11(3):191-200.

289. Kelava I, Lancaster MA. Dishing out mini-brains: Current progress and future prospects in brain organoid research. Dev Biol. 2016.

290. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008;26(3):317-25.

291. Ehrlich M, Mozafari S, Glatza M, Starost L, Velychko S, Hallmann A-L, et al. Rapid and efficient generation of oligodendrocytes from human induced pluripotent stem cells using transcription factors. Proceedings of the National Academy of Sciences. 2017;114(11):E2243.

292. Hirokawa N, Funakoshi T, Sato-Harada R, Kanai Y. Selective stabilization of tau in axons and microtubule-associated protein 2C in cell bodies and dendrites contributes to polarized localization of cytoskeletal proteins in mature neurons. J Cell Biol. 1996;132(4):667-79.

293. Yoo K-S, Lee K, Oh J-Y, Lee H, Park H, Park YS, et al. Postsynaptic density protein 95 (PSD-95) is transported by KIF5 to dendritic regions. Molecular Brain. 2019;12(1):97.

294. Wiedenmann B, Franke WW. Identification and localization of synaptophysin, an integral membrane glycoprotein of M_r 38,000 characteristic of presynaptic vesicles. Cell. 1985;41(3):1017-28.

295. de Jong JO, Llapashtica C, Strauss K, Provenzano F, Sun Y, Cortese GP, et al. Cortical Overgrowth in a Preclinical Forebrain Organoid Model of

CNTNAP2-Associated Autism Spectrum Disorder. bioRxiv. 2019:739391.

296. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816-21.

297. Bassett AR. Editing the genome of hiPSC with CRISPR/Cas9: disease models. Mamm Genome. 2017;28(7-8):348-64.

298. Kan Y, Ruis B, Takasugi T, Hendrickson EA. Mechanisms of precise genome editing using oligonucleotide donors. Genome Res. 2017;27(7):1099-111.

299. Davis L, Maizels N. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. Cell Rep. 2016;17(7):1872-81.

300. Richardson CD, Kazane KR, Feng SJ, Zelin E, Bray NL, Schafer AJ, et al. CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. Nat Genet. 2018;50(8):1132-9.

301. Antoine MW, Langberg T, Schnepel P, Feldman DE. Increased Excitation-Inhibition Ratio Stabilizes Synapse and Circuit Excitability in Four Autism Mouse Models. Neuron. 2019;101(4):648-61.e4.

302. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods. 2009;6(5):377-82.

303. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015;347(6226):1138-42.

304. Ramsköld D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30(8):777-82.

305. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell. 2010;6(5):468-78.

306. Cuevas-Diaz Duran R, Wei H, Wu JQ. Single-cell RNA-sequencing of the brain. Clin Transl Med. 2017;6(1):20.

307. Xing X, Zhang J, Wu K, Cao B, Li X, Jiang F, et al. Suppression of Akt-mTOR pathway rescued the social behavior in Cntnap2-deficient mice. Scientific Reports. 2019;9(1):3041.

308. Macosko Evan Z, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.

309. McInnes L, ohn HJ, James M. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. 2018.

310. Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, et al. Molecular identity of human outer radial glia during cortical development. Cell. 2015;163(1):55-67.

311. Hanashima C, Fernandes M, Hebert JM, Fishell G. The Role of Foxg1 and Dorsal Midline Signaling in the Generation of Cajal-Retzius Subtypes. The Journal of Neuroscience. 2007;27(41):11103.

312. Liu B, Xiao H, Zhao C. Forced Expression of Foxg1 in the Cortical Hem Leads to the Transformation of Cajal-Retzius Cells into Dentate Granule Neurons. J Dev Biol. 2018;6(3).

313. Subramanian L, Remedios R, Shetty A, Tole S. Signals from the edges: the cortical hem and antihem in telencephalic development. Semin Cell Dev Biol. 2009;20(6):712-8.

314. Li J, Sun L, Peng X-L, Qi S-J, Shen Q. Transcriptome analysis of early stage of neurogenesis reveals regulatory gene network for preplate neuron differentiation and CR cell specification. bioRxiv. 2018:331900.

315. Bielle F, Griveau A, Narboux-Nême N, Vigneau S, Sigrist M, Arber S, et al. Multiple origins of Cajal-Retzius cells at the borders of the developing pallium. Nature Neuroscience. 2005;8(8):1002-12.

316. Chiara F, Badaloni A, Croci L, Yeh ML, Cariboni A, Hoerder-Suabedissen A, et al. Early B-cell factors 2 and 3 (EBF2/3) regulate early migration of Cajal–Retzius cells from the cortical hem. Developmental Biology. 2012;365(1):277-89.

317. Yoshida M, Assimacopoulos S, Jones KR, Grove EA. Massive loss of Cajal-Retzius cells does not disrupt neocortical layer order. Development. 2006;133(3):537.

318. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic acids research. 2013;41(Database issue):D996-D1008.

319. Alonso A, Trujillo CM, Puelles L. Longitudinal developmental analysis of prethalamic eminence derivatives in the chick by mapping of Tbr1 in situ expression. Brain Struct Funct. 2020;225(2):481-510.

320. Renner M, Lancaster MA, Bian S, Choi H, Ku T, Peer A, et al. Self-organized developmental patterning and differentiation in cerebral organoids. Embo j. 2017;36(10):1316-29.

321. Caballero IM, Manuel MN, Molinek M, Quintana-Urzainqui I, Mi D, Shimogori T, et al. Cell-autonomous repression of Shh by transcription factor Pax6 regulates diencephalic patterning by controlling the central diencephalic organizer. Cell reports. 2014;8(5):1405-18.

322. Newman EA, Wu D, Taketo MM, Zhang J, Blackshaw S. Canonical Wnt signaling regulates patterning, differentiation and nucleogenesis in mouse hypothalamus and prethalamus. Developmental Biology. 2018;442(2):236-48.

323. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. Genome Biology. 2016;17(1):29.

324. de Lores Arnaiz GR, Ordieres MGL. Brain Na(+), K(+)-ATPase Activity In Aging and Disease. International journal of biomedical science : IJBS. 2014;10(2):85-102.

325. Guillaume D, Grisar T, Delgado-Escueta AV, Bureau-Heeren M, Laschet J. Phosphorylation of brain (Na+,K+)-ATPase alpha catalytic subunits in normal and epileptic cerebral cortex: I. The audiogenic mice and the cat with a freeze lesion. J Neurosci Res. 1991;29(2):207-17.

326. Bagrov AY, Bagrov YY, Fedorova OV, Kashkin VA, Patkina NA, Zvartau EE. Endogenous digitalis-like ligands of the sodium pump: possible involvement in mood control and ethanol addiction. Eur Neuropsychopharmacol. 2002;12(1):1-12.

327. Tochigi M, Iwamoto K, Bundo M, Sasaki T, Kato N, Kato T. Gene expression profiling of major depression and suicide in the prefrontal cortex of postmortem brains. Neurosci Res. 2008;60(2):184-91.

328. Liguri G, Taddei N, Nassi P, Latorraca S, Nediani C, Sorbi S. Changes in Na+,K(+)-ATPase, Ca2(+)-ATPase and some soluble enzymes related to energy metabolism in brains of patients with Alzheimer's disease. Neurosci Lett. 1990;112(2-3):338-42.

329. Binder EB, Kinkead B, Owens MJ, Nemeroff CB. The role of neurotensin in the pathophysiology of schizophrenia and the mechanism of action of antipsychotic drugs. Biol Psychiatry. 2001;50(11):856-72.

330. Boules MM, Fredrickson P, Muehlmann AM, Richelson E. Elucidating the role of neurotensin in the pathophysiology and management of major mental disorders. Behav Sci (Basel). 2014;4(2):125-53.

331. Debonnel G. Current hypotheses on sigma receptors and their physiological role: possible implications in psychiatry. Journal of psychiatry & neuroscience : JPN. 1993;18(4):157-72.

332. Dyment DA, Terhal PA, Rustad CF, Tveten K, Griffith C, Jayakar P, et al. De novo substitutions of TRPM3 cause intellectual disability and epilepsy. Eur J Hum Genet. 2019;27(10):1611-8.

333. Knaus P, Marquèze-Pouey B, Scherer H, Betz H. Synaptoporin, a novel putative channel protein of synaptic vesicles. Neuron. 1990;5(4):453-62.

334. Rogers JH. Calretinin: a gene for a novel calcium-binding protein expressed principally in neurons. J Cell Biol. 1987;105(3):1343-53.

335. Camillo D, Levelt CN, Heimel JA. Lack of functional specialization of neurons in the mouse primary visual cortex that have expressed calretinin. Front Neuroanat. 2014;8:89.

336. González-Gómez M, Meyer G. Dynamic expression of calretinin in embryonic and early fetal human cortex. Front Neuroanat. 2014;8:41.

337. Lukas W, Jones KA. Cortical neurons containing calretinin are selectively resistant to calcium overload and excitotoxicity in vitro. Neuroscience. 1994;61(2):307-16.

338. Fujino T, Wu Z, Lin WC, Phillips MA, Nedivi E. cpg15 and cpg15-2 constitute a family of activity-regulated ligands expressed differentially in the nervous system to promote neurite growth and neuronal survival. J Comp Neurol. 2008;507(5):1831-45.

339. Cantallops I, Haas K, Cline HT. Postsynaptic CPG15 promotes synaptic maturation and presynaptic axon arbor elaboration in vivo. Nat Neurosci. 2000;3(10):1004-11.

340. Javaherian A, Cline HT. Coordinated motor neuron axon growth and neuromuscular synaptogenesis are promoted by CPG15 in vivo. Neuron. 2005;45(4):505-12.

341. Lu J-M, Liu D-D, Li Z-Y, Ling C, Mei Y-A. Neuritin Enhances Synaptic Transmission in Medial Prefrontal Cortex in Mice by Increasing CaV3.3 Surface Expression. Cerebral Cortex. 2017;27(7):3842-55.

342. Fatjó-Vilas M, Prats C, Pomarol-Clotet E, Lázaro L, Moreno C, González-Ortega I, et al. Involvement of NRN1 gene in schizophrenia-spectrum and bipolar disorders and its impact on age at onset and cognitive functioning. World J Biol Psychiatry. 2016;17(2):129-39.

343. Chandler D, Dragović M, Cooper M, Badcock JC, Mullin BH, Faulkner D, et al. Impact of Neuritin 1 (NRN1) polymorphisms on fluid intelligence in schizophrenia. Am J Med Genet B Neuropsychiatr Genet. 2010;153b(2):428-37.

344. Cheng X, Wang H, Zhang X, Zhao S, Zhou Z, Mu X, et al. The Role of SDF-1/CXCR4/CXCR7 in Neuronal Regeneration after Cerebral Ischemia. Frontiers in Neuroscience. 2017;11:590.

345. Bagri A, Gurney T, He X, Zou YR, Littman DR, Tessier-Lavigne M, et al. The chemokine SDF1 regulates migration of dentate granule cells. Development. 2002;129(18):4249-60.

346. Opatz J, Küry P, Schiwy N, Järve A, Estrada V, Brazda N, et al. SDF-1 stimulates neurite growth on inhibitory CNS myelin. Mol Cell Neurosci. 2009;40(2):293-300.

347. Lieberam I, Agalliu D, Nagasawa T, Ericson J, Jessell TM. A Cxcl12-Cxcr4 Chemokine Signaling Pathway Defines the Initial Trajectory of Mammalian Motor Axons. Neuron. 2005;47(5):667-79.

348. Zhou Z, Liu T, Sun X, Mu X, Zhu G, Xiao T, et al. CXCR4 antagonist AMD3100 reverses the neurogenesis promoted by enriched environment and suppresses long-term seizure activity in adult rats of temporal lobe epilepsy. Behav Brain Res. 2017;322(Pt A):83-91.

349. Bonham LW, Karch CM, Fan CC, Tan C, Geier EG, Wang Y, et al. CXCR4 involvement in neurodegenerative diseases. Translational Psychiatry. 2018;8(1):73.

350. Hebebrand M, Hüffmeier U, Trollmann R, Hehr U, Uebe S, Ekici AB, et al. The mutational and phenotypic spectrum of TUBA1A-associated tubulinopathy. Orphanet Journal of Rare Diseases. 2019;14(1):38.

351. Poirier K, Keays DA, Francis F, Saillour Y, Bahi N, Manouvrier S, et al. Large spectrum of lissencephaly and pachygyria phenotypes resulting from de novo missense mutations in tubulin alpha 1A (TUBA1A). Hum Mutat. 2007;28(11):1055-64.

352. Bamba Y, Shofuda T, Kato M, Pooh RK, Tateishi Y, Takanashi J, et al. In vitro characterization of neurite extension using induced pluripotent stem cells derived from lissencephaly patients with TUBA1A missense mutations. Mol Brain. 2016;9(1):70.

353. Grimm J, Sachs M, Britsch S, Di Cesare S, Schwarz-Romond T, Alitalo K, et al. Novel p62dok family members, dok-4 and dok-5, are substrates of the c-Ret receptor tyrosine kinase and mediate neuronal differentiation. Journal of Cell Biology. 2001;154(2):345-54.

354. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;47(6):582-8.

355. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216-21.

356. Mikhail FM, Lose EJ, Robin NH, Descartes MD, Rutledge KD, Rutledge SL, et al. Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. Am J Med Genet A. 2011;155a(10):2386-96.

357. Schanze I, Bunt J, Lim JWC, Schanze D, Dean RJ, Alders M, et al. NFIB Haploinsufficiency Is Associated with Intellectual Disability and Macrocephaly. Am J Hum Genet. 2018;103(5):752-68.

358. Hickey SL, Berto S, Konopka G. Chromatin Decondensation by FOXP2 Promotes Human Neuron Maturation and Expression of Neurodevelopmental Disease Genes. Cell Rep. 2019;27(6):1699-711.e9.

359. Wang W, Mullikin-Kilpatrick D, Crandall JE, Gronostajski RM, Litwack ED, Kilpatrick DL. Nuclear factor I coordinates multiple phases of cerebellar granule cell development via regulation of cell adhesion molecules. J Neurosci. 2007;27(23):6115-27.

360. Xin Y, Li Z, Zheng H, Ho J, Chan MTV, Wu WKK. Neuro-oncological ventral antigen 1 (NOVA1): Implications in neurological diseases and cancers. Cell Proliferation. 2017;50(4):e12348.

361. Fryssira H, Tsoutsou E, Psoni S, Amenta S, Liehr T, Anastasakis E, et al. Partial monosomy14q involving FOXG1 and NOVA1 in an infant with microcephaly, seizures and severe developmental delay. Mol Cytogenet. 92016. p. 55.

362. Li H, Sun C, Wang Y, Gao Y, Liu Y, Li X, et al. Dynamic expression pattern of neuro-oncological ventral antigen 1 (Nova1) in the rat brain after focal cerebral ischemia/reperfusion insults. J Histochem Cytochem. 2013;61(1):45-54.

363. Lee JE, Hollenberg SM, Snider L, Turner DL, Lipnick N, Weintraub H. Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. Science. 1995;268(5212):836.

364. Sega AG, Mis EK, Lindstrom K, Mercimek-Andrews S, Ji W, Cho MT, et al. De novo pathogenic variants in neuronal differentiation factor 2 (NEUROD2) cause a form of early infantile epileptic encephalopathy. Journal of Medical Genetics. 2019;56(2):113.

365. Runge K, Mathieu R, Bugeon S, Lafi S, Beurrier C, Sahu S, et al. Disruption of the transcription factor NEUROD2 causes an autism syndrome via cell-autonomous defects in cortical projection neurons. bioRxiv. 2020:296889.

366. Halgren C, Bache I, Bak M, Myatt MW, Anderson CM, Brøndum-Nielsen K, et al. Haploinsufficiency of CELF4 at 18q12.2 is associated with developmental and behavioral disorders, seizures, eye manifestations, and obesity. Eur J Hum Genet. 2012;20(12):1315-9.

367. Yang Y, Mahaffey CL, Bérubé N, Maddatu TP, Cox GA, Frankel WN. Complex seizure disorder caused by Brunol4 deficiency in mice. PLoS Genet. 2007;3(7):e124.

368. Wagnon JL, Mahaffey CL, Sun W, Yang Y, Chao HT, Frankel WN. Etiology of a genetically complex seizure disorder in Celf4 mutant mice. Genes Brain Behav. 2011;10(7):765-77.

369. Wagnon JL, Briese M, Sun W, Mahaffey CL, Curk T, Rot G, et al. CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. PLoS Genet. 2012;8(11):e1003067.

370. Fitzky BU, Witsch-Baumgartner M, Erdel M, Lee JN, Paik YK, Glossmann H, et al. Mutations in the Delta7-sterol reductase gene in patients with the Smith-Lemli-Opitz syndrome. Proc Natl Acad Sci U S A. 1998;95(14):8181-6.

371. Coman D, Vissers LELM, Riley LG, Kwint MP, Hauck R, Koster J, et al. Squalene Synthase Deficiency: Clinical, Biochemical, and Molecular Characterization of a Defect in Cholesterol Biosynthesis. American journal of human genetics. 2018;103(1):125-30.

372. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science. 2017;358(6368):1318-23.

373. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-d13.

374. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7.

375. Penagarikano O, Lazaro MT, Lu XH, Gordon A, Dong H, Lam HA, et al. Exogenous and evoked oxytocin restores social behavior in the Cntnap2 mouse model of autism. Sci Transl Med. 2015;7(271):271ra8.

376. Lauber E, Filice F, Schwaller B. Dysregulation of Parvalbumin Expression in the Cntnap2-/- Mouse Model of Autism Spectrum Disorder. Front Mol Neurosci. 2018;11:262.

377. Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, et al. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat Methods. 2013;10(5):407-9.

378. Rodriguez A, Ehlenberger DB, Dickstein DL, Hof PR, Wearne SL. Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. PLoS One. 2008;3(4):e1997.

379. Sandoe J, Eggan K. Opportunities and challenges of pluripotent stem cell neurodegenerative disease models. Nat Neurosci. 2013;16(7):780-9.

380. Zaslavsky K, Zhang WB, McCready FP, Rodrigues DC, Deneault E, Loo C, et al. SHANK2 mutations associated with autism spectrum disorder cause hyperconnectivity of human neurons. Nat Neurosci. 2019;22(4):556-64.

381. Shcherbo D, Merzlyak EM, Chepurnykh TV, Fradkov AF, Ermakova GV, Solovieva EA, et al. Bright far-red fluorescent protein for whole-body imaging. Nat Methods. 2007;4(9):741-6.

382. Russell JT. Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. Br J Pharmacol. 2011;163(8):1605-25.

383. Bardy C, van den Hurk M, Eames T, Marchand C, Hernandez RV, Kellogg M, et al. Neuronal medium that supports basic synaptic functions and activity of human neurons in vitro. Proceedings of the National Academy of Sciences. 2015;112(20):E2725.

384. Pak C, Danko T, Zhang Y, Aoto J, Anderson G, Maxeiner S, et al. Human Neuropsychiatric Disease Modeling using Conditional Deletion Reveals Synaptic Transmission Defects Caused by Heterozygous Mutations in NRXN1. Cell Stem Cell. 2015;17(3):316-28.

385. Yi F, Danko T, Botelho SC, Patzke C, Pak C, Wernig M, et al. Autism-associated SHANK3 haploinsufficiency causes Ih channelopathy in human neurons. Science. 2016;352(6286):aaf2669.

386. Shcheglovitov A, Shcheglovitova O, Yazawa M, Portmann T, Shu R, Sebastiano V, et al. SHANK3 and IGF1 restore synaptic deficits in neurons from 22q13 deletion syndrome patients. Nature. 2013;503(7475):267-71.

387. Biffi E, Regalia G, Menegon A, Ferrigno G, Pedrocchi A. The Influence of Neuronal Density and Maturation on Network Activity of Hippocampal Cell Cultures: A Methodological Study. PLOS ONE. 2013;8(12):e83899.

388. Rilling JK, Barks SK, Parr LA, Preuss TM, Faber TL, Pagnoni G, et al. A comparison of resting-state brain activity in humans and chimpanzees. Proc Natl Acad Sci U S A. 2007;104(43):17146-51.

389. Ryu H, Inoue F, Whalen S, Williams A, Kircher M, Martin B, et al. Massively parallel dissection of human accelerated regions in human and chimpanzee neural progenitors. bioRxiv; 2018.

390. Oksenberg N, Stevison L, Wall JD, Ahituv N. Function and regulation of AUTS2, a gene implicated in autism and human evolution. PLoS Genet. 2013;9(1):e1003221.

391. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell. 2013;155(7):1521-31.

392. Knörnschild M. Vocal production learning in bats. Current Opinion in Neurobiology. 2014;28:80-5.

393. Allison AC. Protection afforded by sickle-cell trait against subtertian malareal infection. Br Med J. 1954;1(4857):290-4.

394. Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol. 1985;2(2):150-74.

395. Waterson RH, Lander ES, Wilson RK, The Chimpanzee Sequencing and Analysis C. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005;437(7055):69-87.

396. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585-95.

397. Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. Genetics. 1987;116(1):153-9.

398. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics. 2000;155(3):1405-13.

399. Przeworski M. The signature of positive selection at randomly chosen loci. Genetics. 2002;160(3):1179-89.

400. Weir BS, Cockerham CC. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. Evolution. 1984;38(6):1358-70.

401. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet. 2004;74(6):1111-20.

402. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006;4(3):e72.

403. Gao Z, Przeworski M, Sella G. Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. Evolution. 2015;69(2):431-46.

404. Siewert KM, Voight BF. Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. Mol Biol Evol. 2017;34(11):2996-3005.

405. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002;419(6909):832-7.

406. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007;35(Database issue):D88-92.

407. Paparidis Z, Abbasi AA, Malik S, Goode DK, Callaway H, Elgar G, et al. Ultraconserved non-coding sequence element controls a subset of spatiotemporal GLI3 expression. Dev Growth Differ. 2007;49(6):543-53.

408. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature Methods. 2012;9(3):215-6.

409. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045-8.

410. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317-30.

411. Zhou X, Maricque B, Xie M, Li D, Sundaram V, Martin EA, et al. The Human Epigenome Browser at Washington University. Nature Methods. 2011;8(12):989-90.

412. Prescott SL, Srinivasan R, Marchetto MC, Grishina I, Narvaiza I, Selleri L, et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell. 2015;163(1):68-83.

413. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. Nature. 2016;538(7626):523-7.

414. Varki A, Altheide TK. Comparing the human and chimpanzee genomes: searching for needles in a haystack. Genome Res. 2005;15(12):1746-58.

415. Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS, et al. High-resolution comparative analysis of great ape genomes. Science. 2018;360(6393):eaar6343.

416. Hill RS, Walsh CA. Molecular insights into human brain evolution. Nature. 2005;437(7055):64-7.

417. Peles E, Nativ M, Campbell PL, Sakurai T, Martinez R, Lev S, et al. The carbonic anhydrase domain of receptor tyrosine phosphatase beta is a functional ligand for the axonal cell recognition molecule contactin. Cell. 1995;82(2):251-60.

418. Milev P, Maurel P, Häring M, Margolis RK, Margolis RU. TAG-1/axonin-1 is a high-affinity ligand of neurocan, phosphacan/protein-tyrosine phosphatase-zeta/beta, and N-CAM. J Biol Chem. 1996;271(26):15716-23.

419. Zeng L, D'Alessandri L, Kalousek MB, Vaughan L, Pallen CJ. Protein tyrosine phosphatase alpha (PTPalpha) and contactin form a novel neuronal receptor complex linked to the intracellular tyrosine kinase fyn. J Cell Biol. 1999;147(4):707-14.

420. Li W, Shi L, You Y, Gong Y, Yin B, Yuan J, et al. Downstream of tyrosine kinase/docking protein 6, as a novel substrate of tropomyosin-related kinase C receptor, is involved in neurotrophin 3-mediated neurite outgrowth in mouse cortex neurons. BMC Biol. 2010;8:86.

421. Hays SA, Huber KM, Gibson JR. Altered neocortical rhythmic activity states in Fmr1 KO mice are due to enhanced mGluR5 signaling and involve changes in excitatory circuitry. J Neurosci. 2011;31(40):14223-34.

422. Gonçalves JT, Anstey JE, Golshani P, Portera-Cailliau C. Circuit level defects in the developing neocortex of Fragile X mice. Nat Neurosci. 2013;16(7):903-9.

423. Testa-Silva G, Loebel A, Giugliano M, de Kock CP, Mansvelder HD, Meredith RM. Hyperconnectivity and slow synapses during early development of medial prefrontal cortex in a mouse model for mental retardation and autism. Cereb Cortex. 2012;22(6):1333-42.

424. Peça J, Feliciano C, Ting JT, Wang W, Wells MF, Venkatraman TN, et al. Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. Nature. 2011;472(7344):437-42.

425. Gómez-Robles A. Dental evolutionary rates and its implications for the Neanderthal–modern human divergence. Science Advances. 2019;5(5):eaaw1268.

Appendix 1: Nanostring probes

Gene	*Sequence*	*% Similarity with macaque*
ACTA2	attccttcgttactactgctgagcgtgagattgtccgggacatcaaggagaaactgtgttatgtagctctggactttgaaaatgagatggccactgccgc	**99**
ALDOC	gcatggccaagcggctgagccaaattggggtggaaaacacagaggagaaccgccggctgtaccgccaggtcctgttcagtgctgatgaccgtgtgaaaaa	**96**
ANXA2	tgagcgtccagaaatggtgctcaccatgcttccagctaacaggtctagaaaaccagcttgcgaataacagtccccgtggccatccctgtgagggtgacgt	< 95
ARX	tgcactcagcgtggtatggtaaaagtttgtcctcccgtagattcttactgtgttgtagatacggtagggttcctagacaaatatttatgtactcaagccc	**99**
ASCL1	atagtaactcccatcacctctaacacgcacagctgaaagttcttgctcgggtcccttcacctcctcgccctttcttaaagtgcagttcttagccctctag	< 95
AXIN2	cttgtccagcaaaactctgagggccacggcgagtgtgaggtccacggaaactgttgacagtggatacaggtccttcaaggaggagcgatcctgttaatcct	**99**
B3GAT1	tgtccgtgtggcccgtcgccttcgtgggtggcctgcggtacgaggccccacgggtgaacggggcagggaaggtggtcggctggaagacggtgtttgaccc	**99**
BCL11B	gagatgtagcactcatgtcgtcccgagtcaagcggccttttctgtgttgatttcggctttcatattacataagggaaaccttgagtggtggtgctggggg	< 95
BMP7	gcttcgtcaacctcgtggaacatgacaaggaattcttccacccacgctaccaccatcgagagttccggtttgatctttccaagatcccagaaggggaagc	**100**
CALB1	tggagggaagctgtaccgaacggatcttgctcttattctctgtgctggggataactagagttggtggccgcaaccacttgctagtgatacactgtatcta	**99**
CALB2	cgaaccggccgtacgatgagcccaagctccaggaatacacccaaaccatactacggatgtttgacttgaacggggatggcaaattgggcctctcagagat	**99**
CCK	aaaatgtgtctgtaagattgtccagtgcaaccacacacctcaccagaattgtgcaaatggaagacaaaatgttttcttcatctgtgactcctggtctgaa	**95**
CCND1	ttgaacacttcctctccaaaatgccagaggcggaggagaacaaacagatcatccgcaaacacgcgcagaccttcgttgccctctgtgccacagatgtgaa	**98**
CCND2	tgtgaggaacagaagtgcgaagaagaggtcttccctctggccatgaattacctggaccgtttcttggctggggtcccgactccgaagtcccatctgcaac	**99**
CCND3	ggccagccatgtctgcatttcggtggctagtcaagctcctcctccctgcatctgaccagcagcgcctttcccaactctagctgggggtgggccaggctga	**97**
CDH2	ggtcatccctccaatcaacttgccagaaaactccaggggacctttttcctcaagagcttgtcaggatcaggtctgatagagataaaaaacctttcactgcgg	**100**
CHAT	tcattaatttccgccgtctcagtgaggggatctgttcactcagttgagaaagatagtcaaaatggcttccaacgaggacgagcgtttgcctccaattgg	**99**
CLTC	gggtatcaacccagcaaacattggcttcagtaccctgactatggagtctgacaaattcatctgcattagagaaaaagtaggagagcaggcccaggtggta	**100**
CNN1	gtttgagaacaccaaccatacacaggtgcagtccaccctcctggctttggccagcatggcgaagacgaaaggaaacaaggtgaacgtgggagtgaagtac	**96**
CSPG4	atatattcagggtgcagggtgggtaggttgctctggggatgggtttatttaagggagattgcaaggaagctatttaacatggtgctgagctagccaggac	**97**
CUX1	cataaggttcagagcctacaaacagccctggaaaaaactcgaacagaattatttgacctgaaaaccaaatacgatgaagaaactactgcaaaggccgacg	**100**
CUX2	gataacagaatgtccgtgccattgtaaatgttgtagagatgtgggccgtggcccaaccgtcctatatgagatgtagcatggtacagaacaaactgcttac	**96**
CXCL5	agagagctgcgttgcgtttgtttacagaccacgcaaggagttcatcccaaaatgatcagtaatctgcaagtgttcgccataggcccacagtgctccaagg	< 95
DCX	gctggaaggaaccttaagatcacatcatctactcctctactccaaatttctcattcttcaggccaggaaaccgagacacagaggtaaagtaatttcccca	**99**
DDC	gttgccaccctggggaccacaacatgctgctcctttgacaatctcttagaagtcggtcctatctgcaacaaggaagacatatggctgcacgttgatgcag	**97**

DLG4	tgccctgaagaatgcgggtcagacggtcacgatcatcgctcagtataaaccagaagagtacagccgattcgaggccaagatccacgaccttcgggaacag	**98**
DLX1	ttggagggaggggttttatttattgagaaatggacttcgcctgaggctgtttgccaattcagggttctgctgggcgcaaggaacgcactgttcaaacgc	< 95
DLX4	agctccttgcagctgtctctgttcaaaggaactgtgcagatttagaacgaattggagcttgagcttcacaccaccagcttccagagattaaagtttgtac	< 95
DLX5	ccgagcccgaggtgagaatggtgaatggcaaaccaaagaaagttcgtaaacccaggactatttattccagctttcagctggccgcattacagagaaggtt	**99**
DNMT3B	cgaaggcggcccattcgagtcctgtcattgtttgatggcatcgcgacaggctacctagtcctcaaagagttgggcataaaggtaggaaagtacgtcgctt	**98**
E2F1	ccagctccaagccgtggactcttcggagaactttcagatctcccttaagagcaaacaaggcccgatcgatgttttcctgtgccctgaggagaccgtaggt	**97**
EGFR	acgcagttgggcacttttgaagatcattttctcagcctccagaggatgttcaataactgtgaggtggtccttgggaatttggaaattacctatgtgcaga	**99**
EMX1	gaggaggaagggcctgagtccgagcagaagaagaagggctcccatcacatcaaccggtggcgcattgccacgaagcaggccaatggggaggacatcgatg	**99**
EMX2	acattccctttcctaacatcctgaggcttaaaaccctgatgcaaacttctcctttcagtggttggagaaattggccgagttcaaccattcactgcaatgc	**99**
EN1	gcagcatttttgaaaagggagaaagactcggacaggtgctatcgaaaaataagatccattctctattcccagtataaggggacgaaactgcgaactcctta	**98**
EN2	cacacaccaggcgtgtttgagtccacagttctgaaacatgtggctaccttgtctttcaaaagaactcagaatcctccaggatctagaagaaggaagaaag	< 95
EOMES	atcccatgccctggggtattacccagacccaacctttcctgcaatggcagggtggggaggtcgaggttcttaccagaggaagatggcagctggactacca	**99**
ETV1	cacataccaacggcgaggatcacttcagctctggcagtttttggtagctcttctggatgacccttcaaattctcattttattgcctggactggtcgaggc	**100**
FABP7	gggaaatgtgaccaaaccaacggtaattatcagtcaagaaggagacaaagtggtcatcaggactctcagcacattcaagaacacggagattagtttccag	**97**
FEZF2	cagcacgctctgcaggcacaaaattatccacacccaggaaaagccacataaatgcaaccagtgcggcaaagcgttcaaccgcagctccacgctcaacacg	**98**
FOXA1	tgatacattctcaagagttgcttgaccgaaagttacaaggaccccaaccccttttgtcctctctacccacagatggccctggaatcaattcctcaggaat	**97**
FOXA2	ccgccatgggcagcggctcgggcaacatgagcgcgggctccatgaacatgtcgtcgtacgtgggcgctggcatgagccccgtccctggcggggatgtcccc	**96**
FOXG1	ctgacaagtctatctctaagagccgccagatttccatgtgtgcagtattataagttatcatggaactatatggtggacgcagaccttgagaacaacctaa	**100**
FOXP2	tcgactacctcctccaacacttccaaagcatcaccaccaataactcatcattccatagtgaatggacagtcttcagttctaagtgcaagacgagacagct	**98**
GAD1	caaaggaccaacagcctggaagagaagagtcgccttgtgagtgccttcaaggagaggcaatcctccaagaacctgcttttcctgtgaaaacagcgaccggg	**99**
GAD2	tgtatgccatgatgatcgcacgctttaagatgttcccagaagtcaaggagaaaggaatggctgctcttcccaggctcattgccttcacgtctgaacatag	**96**
GAPDH	cactcctccacctttgacgctggggctggcattgccctcaacgaccactttgtcaagctcatttcctggtatgacaacgaatttggctacagcaacaggg	< 95
GAS1	ctgtggcttgggacagatagaagggatggttggggatacttcccaaaactttttccaagtcaacttggtgtagccggttcccccggccacgactctgggca	< 95
GATA6	gacagtggcgactgcgctgacagaacgtgattctcgtgcctttattttgaaagagatgtttttcccaagaggcttgctgaaagagtgagagaagatggaa	**97**
GBX2	gagccctcccggaaccctaagatcgtcgtccccatccctgtccacgtcagcaggttcgctatcagaagtcagcatcagcagctagaacaggcccggccct	**98**
GFAP	aagcagatgaagccaccctggcccgtctggatctggagaggaagattgagtcgctggaggaggagatccggttcttgaggaagatccacgaggaggaggt	**99**
GLI1	tagcccaagccgtgctaaagctccagtgaacacatatggacctggctttggacccaacttgcccaatcacaagtcaggttcctatcccacccccttcacca	**96**

GLI2	ttcaggagcacattctgattccaggtttggtagagctggctcttctactccgtaaagccgagtctgggactggcagcccatccaagtgtatatgaatgaa	**95**
GLI3	gggacatgagttctttgctgacctccctagcggaagaaagcaaattccttgcagttatgcaataggctttaggaaaaaaagactgcaaccaacggaaatc	**96**
GSC	tcggacagctgacggccgcgggacacttgcccgtattacttacctaactcgaaggacttgcacagacagacgatgctactttcttgcacacgcgctgcct	**99**
GUSB	ccgatttcatgactgaacagtcaccgacgagagtgctggggaataaaaaaggggatcttcactcggcagagacaaccaaaaagtgcagcgttccttttgcg	< 95
HOPX	tcctggagtacaacttcaacaaggtcgacaagcacccggattccaccacgctgtgcctcatcgcggccgaggcaggccttttccgaggaggagacccagaa	**98**
HOXA1	cagataattctggaccagagacttggtgcggggttaacaccttcatccagattgggtgccagcatacattttctggtgggccttaacatccctcctgctt	**97**
HOXA2	cccaaagtttcccagtctcgcctttaaccagcaatgagaaaaatctgaaacattttcagcaccagtcacccactgttccaactgcttgtcaacaatggg	**99**
HOXB1	gggaacgagcagaccgcgagctttgcaccggcctatgctgatctcctctccgaggacaaggaaacaccctgcccttcagaacctaacacccccacggccc	**98**
HOXB2	gaattccactttaataagtacctgtgccggccacgccgcgtcgagatcgcggccttgctggacctcaccgaaaggcaggtcaaagtctggtttcagaacc	**95**
HPRT1	tgtgatgaaggagatgggaggccatcacattgtagccctctgtgtgctcaaggggggctataaattctttgctgacctgctggattacatcaaagcactg	**99**
HTR2C	tctacgttctgcgccgacaagctttgatgttactgcacggccacaccgaggaaccgcctggactaagtctggatttcctgaagtgctgcaagaggaatac	**99**
IGFBP7	cccagaaaagcatgaagtaactggctgggtgctggtatctcctctaagtaaggaagatgctggagaatatgagtgccatgcatccaattcccaaggacag	**99**
IRX3	agtcgcttctgtggcaccccgcattcgctgtgaggtttgtttgtccggttgattttgggggggtggagtttcagtgagaataaacgtgtctgcctttgtgt	**99**
IRX5	cctgtgcaaagactctccctatgaattgaagaaaggtatgtccgacatttaacgcgggctgcgtcggtcccggacttttctaatttattaaaaacatggc	**100**
ISL1	cttacaggctaacccagtggaagtacaaagttaccagccaccttggaaagtactgagcgacttcgccttgcagagtgacatagatcagcctgcttttcag	**99**
KLF4	cgagcattttccaggtcggaccacctcgccttacacatgaagaggcatttttaaatcccagacagtggatatgacccacactgccagaagagaattcagt	**100**
LHX2	acctgggcatctcggcctcggagatggtgatgcgcgctcgggacttggtttatcacctcaactgcttcacgtgcaccacgtgtaacaagatgctgaccac	**100**
LHX6	cgtggattatgtggccgtagcatgttacagttcaaacatgtctccactaccctgttaagagcagcctgggaacgtacaggccatcaagactatttattta	**99**
LHX8	ctcctcacagagcaagatgttaaccatccaaaaccagcaaaaagagctcggaccagctttacagcagatcagcttcaggttatgcaagcacaatttgctc	**98**
LHX9	ctcaagcagcttgcccagaaaacaggtctgaccaaaagagttttgcaggtttggttccaaaacgcacgagccaaattcagaaggaacctttgcggcagg	**99**
LIN28A	gccacagggtgtgtgtgtgtttgtaaaactagagttgctaaggataagtttaaagaccaatacccctgtacttaatcctgtgctgtcgagggatggatat	< 95
LMX1A	ctcccaatgagtttgttatgcgggcccagaagagtgtataccacctgagctgcttctgctgctgtgtctgcgagcgacagcttcagaagggtgatgagtt	**99**
LMX1B	gcaggcacatcacctctcctgctgtggcacccttcctctgttaatttggcccaaaagacaatgatttggccacatgaccttagagattcaccctgccctg	< 95
LRP2	tcttcactgattggttccgtcctgctaaaattatgagagcatggagtgacggatctcacctcttgcctgtaataaacactactcttggatggcccaatgg	**97**
MAP2	tactctgtatgctgggattccgaggttccaacacactgttacaaatctgtgggggtttctttcttctgataattctagagcctgttaccatagaaaggc	**99**
MAPT	attgggtccctggacaatatcacccacgtccctggcggaggaaataaaaagattgaaacccacaagctgaccttccgcgagaacgccaaagccaagacag	< 95
MESP1	ctgcctgaggagcccaagtgacaagggacaactgacgccgtctctgtgagcaccgaggctttttggcctcagcaccttcgaagtggttccttggcagact	< 95

MKI67	agcagatgtagagggagaactcttagcgtgcaggaatctaatgccatcagcaggcaaagccatgcacacgcctaaaccatcagtaggtgaagagaaagac	< 95
MYBL2	gaatccagaccttgtcaaggggccatggaccaaagaggaagaccaaaaagtcatcgagctggttaagaagtatggcacaaagcagtggacactgattgcc	**96**
MYC	tcggacaccgaggagaatgtcaagaggcgaacacacaacgtcttggagcgccagaggaggaacgagctaaaacggagctttttgccctgcgtgaccaga	**99**
MYH11	ctgctagaaaaatcacgggcaattcgccaagccagagacgagaggacattccacatcttttactacatgattgctggagccaaggagaagatgagaagtg	**97**
NANOG	ctactccatgaacatgcaacctgaagacgtgtgaagatgagtgaaactgatattactcaatttcagtctggacactggctgaatccttcctctcccctcc	< 95
NES	cagagaatcacaaatcactgaggtctttagaagaacaggaccaagagacattgagaactcttgaaaaagagactcaacagcgacggaggtctctagggga	< 95
NEUROD1	gtgcccagctcaatgccatatttcatgattagaggcacgccagtttcaccatttccgggaaacgaacccactgtgcttacagtgactgtcgtgtttacaa	**99**
NEUROD2	gtgcctcggtgggtgccacctggcgatttccggtgtctggagagagtatttttggtccaaggagtcctcttggctttagctggtgggtgggcggggaga	< 95
NEUROG1	gcccctagacggccttttccttttgcactttctgaactccacaaaacctcctttgtgactggctcagaactgaccccagccaccacttcagtgtgatttag	**99**
NEUROG2	tttgggtatccttcattcagacgggctctgatttactgaaggtgtgatggagcttattgtcaaagccaagggtggcgttttgggggcgcttcttgagacg	**98**
NFIX	aagaactggatctttatctggcttactttgtccacactccggaatccggacaatcagatagttcaaaccagcaaggagatgcggacatcaaaccactgcc	**100**
NGFR	tgaagaaaagtgggccagtgtgggaatgcggcaagaaggaattgacttcgactgtgacctgtggggatttctcccagctctagacaaccctgcaaaggac	< 95
NKX2-1	gccgcctaggctcagcggcgaccgccctccgcgaaaatagtttgtttaatgtgaacttgtagctgtaaaacgctgtcaaaagttggactaaatgcctagt	**96**
NKX2-2	taattattattatggagtcgagttgactctcggctccactagggaggcgccgggaggttgcctgcgtctccttggagtggcagattccacccacccagct	**96**
NKX6-1	ctggcctgtacccctcatcaaggatccattttgttggacaaagacgggaagagaaaacacacgagacccactttttccggacagcagatcttcgccctgg	**99**
NKX6-2	aggacccgcggggtgggggcgaatctatttttgcagaatccgggggcggccccgggtgggcgcgagtcgctttgtatcatcaataaattatttaacgggt	**97**
NOG	acagagaaaagagagacttattctggttgttgctaataatgttaacctgctatttatattccagtgcccttcgcatggcgaagcaggggggaaaagttat	**100**
NPY	agagatatggaaaacgatccagcccagagacactgatttcagacctcttgatgagagaaagcacagaaaatgttcccagaactcggcttgaagaccctgc	**97**
NR4A2	ttcagaagtgcctggctgttgggatggtcaaagaagtggttcgcacagacagtttaaaaggccggagaggtcgtttgcctcgaaaccgaagagcccaca	**100**
NR5A1	ctctgcaagagggggcattgatacatcatcgggaaaaaactttgctccaggcatcactgattccctctcccacccaaggagaacgtttggtacaatcgac	< 95
OLIG1	gggacgttaaagtgaccagagcggatgttcgatggcgcctcggggcagtttggggttctgggtcggttccagcggctttaggcagaaagtgctcgctctc	< 95
OLIG2	tagttggaagccggcgttcggtatcagaagcgctgatggtcatatccaatctcaatatctgggtcaatccacaccctcttagaactgtggccgttcctcc	< 95
OTP	agggacagcacaatgttagggatttttgtcttaaaggaggacaagcattgctaccaaccgcctcatctgagggcccaactgatatgatttgatttatcct	**98**
OTX1	gaaactcaacttcaactcccccgactgtctggactataaggaccaagcctcatggcggttccaggtcttgtgagcccaggaatgaaagaggagaagaaac	**98**
OTX2	ggctggacattccagttttagccaggcattggttaaaagagttagatgggatgatgctcagactcatctgatcaaagttccgagaggcatagaaggaaaa	**100**
PAX3	gtggagaagaaaattgaggaatacaaaagagagaacccgggcatgttcagctgggaaatccgagacaaattactcaaggacgcggtctgtgatcgaaaca	< 95
PAX6	gaacatcctttacccaagagcaaattgaggccctggagaaagagtttgagagaacccattatccagatgtgtttgcccgagaaagactagcagccaaaat	**100**

PAX7	ccacaggctgacttctccatctccccgctgcatggcggcctggactcggccacctccatctcagccagctgcagccagcgggccgactccatcaagccag	**100**
PDGFRA	tagtgcttggtcgggtcttggggtctggagcgtttgggaaggtggttgaaggaacagcctatggattaagccggtcccaacctgtcatgaaagttgcagt	**100**
PGK1	gcaagaagtatgctgaggctgtcactcgggctaagcagattgtgtggaatggtcctgtgggggtatttgaatgggaagcttttgcccggggaaccaaagc	< 95
PLK1	gcttggctgccagtacctgcaccgaaaccgagttattcatcgagacctcaagctgggcaacctttttcctgaatgaagatctggaggtgaaaataggggat	**98**
POU3F2	gggcacggagctgcttcgggtgcatcacgctgctcgttcctgaggtatgggaactggcctttagtgaagctatccagagcagggcaaatagccactggta	**98**
POU4F1	cacacattcacacagtggtaacagactgccagtgttcatcctgaaatgtctcacggattgatctacctgtctatgtatgtctgctgagctttctccttgg	< 95
POU4F2	tccagtaaatgtgaatctcgacaaatcgaggactgaagagggagcgaacgagcgaacaactgagcccaagccggtgagaatgtgaaacagtttctcaaag	**95**
POU5F1	aagttcttcattcactaaggaaggaattgggaacacaaagggtgggggcaggggagtttggggcaactggttggagggaaggtgaagttcaatgatgctc	**100**
PPIA	agcactggagagaaaggatttggttataagggttcctgctttcacagaattattccagggtttatgtgtcagggtggtgacttcacacgccataatggca	< 95
PPP1R1B	accggcttctaccagggtccaggactaaggcgtttttctccatagcctcaacattttgggaatcttcccttaatcacccttgctcctcctgggtgcctgg	**95**
PTCH1	tggtcatcctgatcgcttctgttggcataggagtggagttcaccgttcacgttgctttggcctttctgacggccatcggcgacaagaaccgcagggctgt	**97**
PVALB	tgaggacatcaagaaggcggtgggagcctttagcgctaccgactccttcgaccacaaaaagttcttccaaatggtcggcctgaagaaaaagagtgcggat	**98**
RELN	ggcaaccccacctactacgttccgggacaagaataccatgtgacaatttcaacaagcacctttttgacggcttgctggtgacaggactatacacatcta	**98**
RORB	cctggatgatgagaccttggcaaagttaatagccaagataccaaccatcacggcagtttgcaacttgcacggggagaagctgcaggtatttaagcaatct	**100**
RPLP1	ctcagctgccgccaaggtgctcggtccttccgaggaagctaaggctgcgttggggtgaggccctcacttcatccggcgactagcaccgcgtccggcagcg	< 95
RPS15A	agctgggaaaattgttgtgaacctcacaggcaggctaaacaagtgtggggtgatcagccccagatttgacgtgcaactcaaagacctggaaaaatggcag	< 95
RPS9	aagcaggtggtgaacatcccgtccttcattgtccgcctggattcccagaagcacatcgacttctctctgcgctctccctacgggggtggccgcccgggcc	**95**
S100B	gagacggcgaatgtgacttccaggaattcatggcctttgttgccatggttactactgcctgccacgagttctttgaacatgagtgagattagaaagcagc	< 95
SATB2	ggccaaggccgtgggaggtttgatgattcctgtcttttgtgtcgtggagcagttggacggctctcttgaatatgacaacagagaagaacacgccgagttt	**98**
SHH	tctgcactacgagggccgcgcagtggacatcaccacgtctgaccgcgaccgcagcaagtacggcatgctggcccgcctggcggtggaggccggcttcgac	< 95
SIX3	cagcaccaggccattggaccgagcggcatgcgctcgctggccgagcccggctgccccacgcacggctcggcagagtcgccgtccacggcggccagcccga	**100**
SIX6	cagcggtgaggcctgacccagcaccacgttcttcttgctttgctttttcctaaggattttgctgcaaagtctccttcggaaccccgaactgcaagctgagc	**96**
SLC17A7	tcggctactcgcactccaagggcgtggccatctccttcctggtcctagccgtgggcttcagcggcttcgccatctctgggttcaacgtgaaccacctgga	**100**
SLC6A3	ccattttgcagccggcacgtccatcctctttggagtgctcatcgaagccatcggagtggcctggttctatggtgttgggcagttcagcgacgacatccag	**97**
SMO	aacgagaccatgctgcgcctgggcatttttggcttcctggcctttggctttgtgctcattaccttcagctgccacttctacgacttcttcaaccaggctg	**98**
SNAI2	gcgttttccagaccctggttgcttcaaggacacattagaactcacacgggggagaagccttttttcttgccctcactgcaacagagcatttgcagacaggt	**98**
SOX1	tatttatcacctacggaggaagcggaaagcgttttctttgctcgaggggacaaaaaagtcaaaacgaggcgagaggcgaagcccacttttgtataccggc	< 95

SOX10	ggccgtgtctcccactcaggggctgagagtagctttgaggagcctcattggggagtgggggttcgagggacttagtggagttctcatcccttcaatgcc	**98**
SOX11	ctaagcattgacagaatatcttaaaatggtaacctgggggtggcgggtgggtgctgtgtgcacggcagcctagccagtgggatcctgctgtttattata	< 95
SOX17	agccacggggccatttcctcggtggtgtccgacgccagctccgcggtatattactgcaactatcctgacgtgtgacaggtccctgatccgccccagcctg	**97**
SOX2	cttaagcctttccaaaaaataataataacaatcatcggcggcggcaggatcggccagaggaggagggaagcgctttttttgatcctgattccagtttgcc	**97**
SOX4	gttcacggtcaaactgaaatggatttgcacgttggggagctggcggcggcggctgctgggcctccgccttctttttctacgtgaaatcagtgaggtgagac	< 95
SOX5	tagccatgcaatgatggatttcaatctgagtggagattctgatggaagtgctggagtctcagagtcaagaatttataggggaatcccgagggcgtggtagc	**97**
SP8	ggtgggggtgacttttggcagcccgagacccagagtttgtagagtttagatctcccaatgtcagtctgcctctttttctttgttccctgaggtggaaccg	< 95
SST	agctgctgtctgaacccaaccagacggagaatgatgccctggaacctgaagatctgtcccaggctgctgagcaggatgaaatgaggcttgagctgcagag	**97**
SULF1	tagaacgaggcattttgaatcagctacacgtacaactaatggagctcagaagctgtcaaggatataagcagtgcaacccaagacctaagaatcttgatgt	**97**
SYP	tagtgcctgtgatcgtgtgttgccattttgtctggctgtggcccctccttctcccctccagacccctaccctttcccaaaccctcggtattgttcaaag	**100**
T	tttttgtcgtggcagccagtggtgactggattgacctactaggtacccagtggcagtctcaggttaagaaggaaatgcagcctcagtaacttccttttca	< 95
TBP	acagtgaatcttggttgtaaacttgacctaaagaccattgcacttcgtgcccgaaacgccgaatataatcccaagcggttgctgcggtaatcatgagga	**97**
TBR1	gccgtctgcagcgaataagtgcaggtctccgagcgtgattttaacctttttgcacagcagtctctgcaattagctcaccgaccttcaactttgctgtaa	< 95
TFAP2B	ctcacttctggcagcccagctcctttcttgggctccatcttagcattatcatgaaataagatctggaatccattgtctgcacctccgcaaaagcagtgag	**99**
TH	agctgattgctgagatcgccttccagtacaggcacggcgacccgattccccgtgtggagtacaccgccgaggagattgccacctggaaggaggtctacac	**100**
TNC	tgtgtgccacgatggctttgcaggcgatgactgcaacaagcctctgtgtctcaacaattgctacaaccgtggacgatgcggtggagaatgagtgcgtgtgt	**97**
TNFRSF19	ttgttcagccaccagtgatgccatctgcggggactgcttgccaggattttataggaagacgaaacttgtcggctttcaagacatggagtgtgtgccttgt	**97**
TPH2	cacaatcgagtttggccctttgcaagcaagaagggcaactgcgggcatatggagcaggactcctttcctccattggagaattaaagcacgccctttctgac	**98**
TUBB3	gccgccctcctgcagtatttatggcctcgtcctccccacctaggccacgtgtgagctgctcctgtctctgtcttattgcagctccaggcctgacgtttta	< 95
UXT	actccatgaatatcaaagcccatatccacatgttgctagaggggcttagagaaactacaaggcctgcagaatttcccagagaagcctcaccattgacttct	**98**
VIM	gaggagatgcttcagagagaggaagccgaaaacaccctgcaatctttcagacaggatgttgacaatgcgtctctggcacgtcttgaccttgaacgcaaag	**99**
VSX2	tgtggcatgtggtgatgtacgctgcgtgccatgagtccatgtcctatgcctcacaaatgctgtggttcactgcactgttcaggagtccaaaccttctacc	< 95
WLS	tcccagtggaatggttttccatcgggtttgactggacctggatgctgctgtttggtgacatccgacagggcatcttctatgcgatgcttctgtccttctg	**98**

Appendix 2: *Multiple species alignments of the CNTNAP2 HARs*

HACNS_97:

- Intron 1, 109bp
- **Hg38:** chr7:146290445-146290553
- **CHIMP2.1.4:** chr7:147595317-147595425
- **gorGor4:** chr7:145869346-145869454
- **Macaca_fascicularis_5.0:** chr3:179295535-179295643

```
homo_sapiens/1-112         GAAGCTTATCCGCCAATTCACAATTCTGAAGCACAGCAAAGCTGTTTCTGTATTG-GTGT
pan_troglodytes2/1-112     AAAGCTTATCTGCCAATTCACAATTCTCAAGCACAGCAAAGCTGTTTCTATATTG-GTGT
gorilla_gorilla/1-112      AAAGCTTATCCGCCAATTCACAATTCTCAAGCACAGCAAAGCTGTTTCTATATTG-GTGT
macaca_fascicularis/1-112 AAAGCTTATCTGCCAATTCACAATTCTCTAGCACAGCAAAGCTGTTTCTATATTG-GTGT
                           ********* ***************  ******************** ***** ****

homo_sapiens/1-112         TAATTATTTATTAA--TGCAGTGAAAACACTTTAAATTCCTATTAAATTGTC
pan_troglodytes2/1-112     TAATTATTTATTAA--TGCAGTGAAAACACTTTAAATTCCAATTAAATTGTC
gorilla_gorilla/1-112      TAATTATTTATTAA--TGCAGTGAAAACACTTTAAATTCCAATTAAATTGTC
macaca_fascicularis/1-112 TAATTATTTATTAA--TGCAGTGAAAACACTTTAAATACCAATTAAATTGTC
                           **************  ********************** ** ***********
```

HACNS_116:

- Intron 1, 196bp
- **Hg38:** chr7:146214973-146215168
- **CHIMP2.1.4:** chr7:147520051-147520246
- **Macaca_fascicularis_5.0:** chr3:179225513-179225708

```
homo_sapiens/1-196         GGAACATTTGGCTTAAAGTTTCAGAAAATCTGCTCCGTTTGATAATGTCAACTCATCACA
pan_troglodytes2/1-196     GGAACATTTGGCTTAAAGTTTCAGAAAATCTGCTCAGTTTGATAATGTCAACTCATCACA
macaca_fascicularis/1-196 GGAACATTTGGCTTAAAGTTTCAGAAAATCTGCTCAGTTTGATAATGTCAACTCATCACA
                           ********************************** ***************** ******
homo_sapiens/1-196         TTTCAGTTAAGCAGACATCAAATCAAATGACTCATGTCACAGATACATAAATATTACGTA
pan_troglodytes2/1-196     TTTCAGTTAAGCAGACATCAAATCAAATGATGCATGTCACAGATACATAAATATTACGTA
macaca_fascicularis/1-196 TTTCAGTTAAGCAGGCATCAAATCAAATGATGCATGTCACAGATACATAAATATTATGTA
                           ************** ***************  ************************ ***
homo_sapiens/1-196         CATGGAAAGGAAATTTTTAGCTATAAATGCATAAAATCTTCTGGTTTGTAAACCTGAACT
pan_troglodytes2/1-196     CATGGAAAGCAAATTTTTAGCTATAAATGCATAGAATCTTCTGGTTTGTAAACCTGAACT
macaca_fascicularis/1-196 CATGGAAAGCAAATTTTTAGCTATAAATGCATAGAATCTTCTGGTTTGTAAACCTGAACT
                           ********* ********************** *************************
homo_sapiens/1-196         ACATTTTAAAATTTTG
pan_troglodytes2/1-196     ACATTTTAAAATTTTG
macaca_fascicularis/1-196 ACATTTTAAAATTTTG
                           ****************
```

HACNS_590:

- Intron 13, 300bp
- **Hg38:** chr7:147859118-147859418
- **CHIMP2.1.4:** chr7:149154567-149154867
- **gorGor4:** chr7:147450614-147450914
- **Macaca_fascicularis_5.0:** chr3:180839557-180839857

```
homo_sapiens/1-308         AAGATGATTTTTT-CAAGCATATGTTACCAAGAAGTAGG----GAAGTTTCAGCATTAAC
pan_troglodytes/1-308      AAGATGATTTTTT-CAACCATATGTTACCAAGAAGTAGG----GAAGTTTCAGCATTCAC
gorilla_gorilla/1-308      AAGATGATTTTTT-CAACCACATGTTACCAAGAAGTAGG----GAAGTTTCAGCATTAAC
macaca_fascicularis/1-308  AAGATGATTTTTT-CAACCATATGTTACCAAGAAGTAGG----GAAGTTTCAGCATTAAC
                           ************* *** ** ******************    ************** **
homo_sapiens/1-308         A--ATACATAGCTTCGTAACAATTAGCCATCTGTTTATAATGCTGTTAGGGATCGACAGC
pan_troglodytes/1-308      A--ATACATAGCTTCGTAACAATTAGCCATCTGTTTATAATGCTGTTAGGGATCGACAGC
gorilla_gorilla/1-308      A--ATACATAGCTTCGTAACAATTAGCCATCTGTTTATAATGCTGTTAGGGATCGACAGC
macaca_fascicularis/1-308  A--ATACAGAGCTTCGTAACAATTAGCCATCTGTTTATAATGATGTTAGGGATTGACAGC
                           *  ***** ********************************** ********** ******
homo_sapiens/1-308         ATCTCAATGGAAGCAGGGAAAACAACAGAAATATCATATCTGCCAAGTTCTAGTCATCTG
pan_troglodytes/1-308      ATCTCAATGGAAGCAGGGAAAACAACAGAAATATCATATCTGCCAAGTTCTAGTCATTTG
gorilla_gorilla/1-308      ATCTCAATGGAAGCAGGGAAAACAACAGAAATATCATATCTGCCAAGTTCTAGTCATTTG
macaca_fascicularis/1-308  ATCTCAATGGAAGCAGGGAAAACAACAGAAATATTATATCTGCCAAGTTCTAGTCATTCG
                           ********************************** ********************  *
```

```
homo_sapiens/1-308         TTATGTTACATAGTAATTTGTCATCTCATGAGATCACCAAGGAGAGAAAAAAGCCATGCC
pan_troglodytes/1-308      TTATCTTACATAGTAATTTGTTGTCTCATGAGATCACCAAGGAGAGGAAAAAGCCATGCC
gorilla_gorilla/1-308      TTATCTTACATAGTAATTCGTCGTCTCATGAGATCACCAAGCAGAGAAAAAAGCCATGCC
macaca_fascicularis/1-308  TTATCTTACATAGTAATTTGCTGTCTCATGAGATCACTAAGGAGAGAAAAACACCATGCC
                           **** ************* *   ************** *** **** ****  *******
homo_sapiens/1-308         GTGCTACAAAATAAACTCTGTAAAGCAGGTAGTTTAATAGCAAATTAATGATGTGCTGAC
pan_troglodytes/1-308      GTGCTACAAAATAAACTCTGTAAAGCAGGTAGTTTAATAGCAAATTAATGATGTGCTGAC
```

```
gorilla_gorilla/1-308     GTGCTACAAAATAAACTCTGTAAAGCAGGTAGTTTAATAGCAAATTAATGATGTGCTGAC
macaca_fascicularis/1-308 GTGCTACAAAATAAACTTTGTAAAGCAGGTAGTTTAATAGCAAATTAATGATGTGCTGAC
                          ****************** ****************************************

homo_sapiens/1-308        AAGATCAA
pan_troglodytes/1-308     AAGATCAA
gorilla_gorilla/1-308     AAGATCAA
macaca_fascicularis/1-308 AAGATCAA
                          ********
```

HACNS_884:

- Intron 1, 347bp
- **Hg38:** chr7:146654063-146654409
- **CHIMP2.1.4:** chr7:147957082-147957428
- **Macaca_fascicularis_5.0:** chr3:179642424-179642770

```
homo_sapiens/1-353         TGTAAAGCTGAGGCCCAGGAGAAGCTCTTTTTC-----TTACCCATTTGGTCTTTAATGT
pan_troglodytes/1-353      TGTAAAGCTGAGGCCCAGGAGAAACTCTTTTTC-----TTACCCATTTGGTCTTTAATGT
macaca_fascicularis/1-353  TATAAAGCTGAGGCCCAGGGGGAGGTCTTTTTC-----CCACCCATTTGGTTTTTAATGT
                           * ***************** * *  ********       *********** ********
homo_sapiens/1-353         TGCTTGCGAATGCATGGGGGTAGAACTCTAATAACAATGATGGACACGATGTCCTTGCAC
pan_troglodytes/1-353      TGCTTGCGAATGCATGGGGGTAGAACTCTAATAACAATAATGGACACAATGTCCTTGCAC
macaca_fascicularis/1-353  TGCTTGTGAATGCATGGGGGTAGAACTCTAATAACAATAATGGACACGATGTCCTTGCAC
                           ****** ****************************** ******** ************
homo_sapiens/1-353         ATAAACATTTGGCTCCTCACAGCATGAGATGTCTGGTCAGCCACAGAGGATAAAATAATG
pan_troglodytes/1-353      ATAAACATTTGGCTCCTCACAGCATGAGATGTCTTGTCAGCCACAGAGGATAAAATAATG
macaca_fascicularis/1-353  ATAAACATTTGGCTCCTCACAGCGTGAGATGTCTTGTCAGCCACAGAGGATAAAATAATG
                           ************************ ********** *************************

homo_sapiens/1-353         GCGTAAAGATTTACTGTCAGGGAGCTGGGCCAGTCTCAGCAGGACAGCGGCTCCAAGCAG
pan_troglodytes/1-353      GCGTAAAGATTTACTGTCAGGGAGCTGGGCCAGTCTCAGCAGGACGGCGGCTCCAAGCAG
macaca_fascicularis/1-353  GCGTAAAGATTTACTGTCAGGGAGCTGGGCCAGTCTCAGCAGGATGGCAGCTCCAAGCAG
                           ******************************************  ** ***********
homo_sapiens/1-353         ATAGATGGC-CTGGTATGCGTAGGTCACCTGTGTTTACACTGATGAATGTGTGTGGTTTC
pan_troglodytes/1-353      ATAGATGGC-CTGGTATACGTAGGTCACCTGTGTTTACACTGACGAATGTGTGTGGTTTC
macaca_fascicularis/1-353  ATAGATGGC-CTGGCATGCGTAGGTCACCTGTGTTTACACTGAGGAATGTGTGTGGTTTC
                           ********* **** ** ************************ ****************
homo_sapiens/1-353         CCTAGTCTGCATCATTATATTGAGTTGTGTGGTGTTGGTAGCACTTGTTACTC
pan_troglodytes/1-353      CCTAGTCTGCATCATTATATTGAGTTGTGTGGTGTTTGTAGCACTTGTAACTC
macaca_fascicularis/1-353  CCTAGTCTGCATCATTATATTGAGTTGTGTGGTGTTTGTAGCACTTGTAACTC
                           ************************************ *********** ****
```

HACNS_954:

- Intron 18, 510bp
- **Hg38:** chr7:148173396-148173905
- **CHIMP2.1.4:** chr7:149470105-149470614
- **gorGor4:** chr7:147774737-147775247
- **Macaca_fascicularis_5.0:** chr3:181174268-181174777

```
homo_sapiens/1-511          ATAAGAAGCTTTACATAAGTAGCCTGTTCAGACTCTGATGAATACATGG-AGGGTTGTTT
pan_troglodytes/1-511       ATAAGAAGCTTTACATAAGTAGCCTATTCAGACTCTGATGAATACATGG-AGGATTGTTT
gorilla_gorilla/1-511       ATAAGAAGCTTTACATAAGTAGCCTGTTCAGACTCTGATGAATACATGGGAGGGTTGTTT
macaca_fascicularis/1-511   ATAAGAAGCTTTACATAAGTAGCCTATTCAGACTCTGATGAATACATGG-AGGGTTGTTT
                            ************************* ********************** *** ******
homo_sapiens/1-511          ACACTGGCATTATTCACACAAGATGCCACACACAGAAGGAGGAATTGTTCCATAACATTC
pan_troglodytes/1-511       ACACTGGCATTATTCACACAAGATGCCACACACAGAAGGAGGAATTGTTCCATAACATTC
gorilla_gorilla/1-511       ACACTGGCATTATTCACACAAGATGCCACACACAGAAGGAGGAATTGTTCCATAACATTC
macaca_fascicularis/1-511   ACACTGGCATTATTCACGCAAGATGCCACACACAGAAGGAGGAATTGTTCCATAACATTC
                            ***************** ****************************************
homo_sapiens/1-511          TCTGTACCTCCAGGCTACTGTTTGGCAGCGCTATAATATGAGCATACAAATTAGGCCAGA
pan_troglodytes/1-511       TCTGTACATCCAGGCTACTGTCTGGCAGCGCTATAATATGAGCATACAAATTAGGCCAGA
gorilla_gorilla/1-511       TCTGTACATCCAGGCTACTGTCTGGCAGCGCTATAATATGAGCATACAAATTAGGCCAGA
macaca_fascicularis/1-511   TCTGTACATCCAGGCTACTGTCTGGCAGCGCTATAATATGAGCATACAAATTAGGCCAGA
                            ******* ************* ************************************

homo_sapiens/1-511          CAGGGTATGAAAAATGAGCATGAGACAACAACAAAGGTGGAAAGGAACACCATTGCAGAA
pan_troglodytes/1-511       CAGGGTATGAAAAATGAGCATGAGACAACAACAAAGGTGGAAAGGAACGCCATTGCAGAA
gorilla_gorilla/1-511       CAGGGTATGAAAAATGAGCATGAGACAACAACAAAGGTGGAAAGGAACACCATTGCAGAA
macaca_fascicularis/1-511   CAGGGTATGAAAAATGAGCATGAGACAACAACAAAGGTGGAAAGGAACACTATAGCGGAA
                            ********************************************* * ** ** ***
homo_sapiens/1-511          TAATACCAAAGCACATTAACAGTGCTATTACAGTGCATTTTTCTGTTCAGAACAACATGG
```

```
pan_troglodytes/1-511     TAATACCATAGCACATTAACAGTGCTATTACAGTGCGTTTTTCTGTTCAGAACAACATGG
gorilla_gorilla/1-511     TAATACCATAGCACATTAACAGTGCTATTACAGTGCATTTTTCTGTTCAGAACAACATGG
macaca_fascicularis/1-511 TAATTCCATACCACACTAACAGCGCTATTACAGTGCATTTTTCTGTTCAGAACAACATGG
                          **** *** * **** ****** ************* **********************

homo_sapiens/1-511        TTCCATGATTTTCCTGCGGCACACAATTGAATGGGCCTGCAGAACACATTAATTTTGCAC
pan_troglodytes/1-511     TTCCATGATTTTCCTGTGGCACACAATGGAATGGGCCTGCAGAACACATTAATTTTGCAC
gorilla_gorilla/1-511     TTCCATGATTTTCCTGTGGCACACAATGGAATGGGCCTGCAGAACACATTAACTTTGCGC
macaca_fascicularis/1-511 TTCCATGATTTTCCTGTGGCACACAATGGAACGGGCTTGCAAAACACATTAATTTTGCAC
                          ***************** ********** *** **** **** ********** ***** *

homo_sapiens/1-511        TTGTAGAAATGACAGTCTTGCAATTACCATTCATTGTGCTTGATGCTAATTTAATGCAGT
pan_troglodytes/1-511     TTGTAGAAATGACAGTCTTGCAATTACCATTCATTGTGCTTGATGCGAATTTAATGCAGT
gorilla_gorilla/1-511     TTGTAGAAATGACAGTCTTGCAATTACCATTCATTGTGCTTGATGCTAATTTAATGCAGT
macaca_fascicularis/1-511 TTGTAGAAATGACAGTCTTGCAATTACCATTCATTGTGCTTGATGCTAATTTAATGCAGT
                          ******************************************** ************

homo_sapiens/1-511        CATTTCCTTGCCAAAGTAATGTCTGAATGCAATCAAGTAAGCACTAATGCCTGCGAGATT
pan_troglodytes/1-511     CGTTTCCTTGCCAAAGTAATGTCTGAATGCAATCAAGTAAGCACTAATGCCTGCAAGATT
gorilla_gorilla/1-511     CATTTCCTTGCCAAAGTAATGTCTGAATGCAATCAAGTAAGCACTAATGCCTGTGAGATT
macaca_fascicularis/1-511 CATTTCCTTGCCAAAGTAATGTCTGAATGCAATCAAGTAAGCACTAATGCCTGAGAGATT
                          * ************************************************  *****

homo_sapiens/1-511        CCACTACATAAGTACCAGTTTAAAGAAAATC
pan_troglodytes/1-511     CCACTACATAAGTACCAATTTAAAGAAAATC
gorilla_gorilla/1-511     CCACTACATAAGTACCAATTTAAAGAAAATC
macaca_fascicularis/1-511 CCAGTATGTAAGTACCAATTTAAAGAAAATC
                          *** **  ********* **************
```

2xHAR.395:

- Intron 1, 24bp
- **Hg38:** chr7:146420328-146420351
- **CHIMP2.1.4:** chr7:147725719-147725742
- **gorGor4:** chr7:146003019-146003042
- **Macaca_fascicularis_5.0:** chr3:179417246-179417269

```
homo_sapiens/1-24         AAGGGTTTCTGTTAATGAACAAAG
pan_troglodytes/1-24      AAGTGCTTCTTTTAATGAACAAAG
pan_troglodytes2/1-24     AAGTGCTTCTTTTAATGAACAAAG
gorilla_gorilla/1-24      AAGTGCTTCTTTTAATGAACAAAG
macaca_fascicularis/1-24 AAGTGCTTCTTTTGATGAACAAAG
                          *** * **** ** **********
```

ANC1209:

- Intron 13, 199bp
- **Hg38:** chr7:147878720-147878918
- **Pan_tro_3.0:** chr7:152387358-152387556
- **gorGor4:** chr7:147470369-147470567
- **Macaca_fascicularis_5.0:** chr3:180856546-180856744

```
homo_sapiens/1-200         AATAAGGCAACATACATCAAATCCTTTTGAGAATATTCAACA-TGAGATAGTTTTCTTTT
pan_troglodytes/1-200      AATAAGGCAACATACATCAAATCCTTTTGAGAATATACAACA-TGAGATAGTTTTCTTTT
gorilla_gorilla/1-200      AATAAGGCAACATACATCAAATCCTTTTGAGAATATACAACA-TGAGATAGTTTTCTTTT
macaca_fascicularis/1-200  AATAAGGCAACATACATCAAATCCTTTTGAGAATATACAACA-AGAGATAGTTTTCTTTT
                           ************************************ *****  ***************
homo_sapiens/1-200         AAGTCTGACAGTGGCAGGCCATAATACTACTGAATAAATTCTTACCAAACTGAAGCTTCT
pan_troglodytes/1-200      AAGTCTGACAGTGGCAGGCCATAATACTACTGAATAAATTCTTACCAAACTGAAGCTTCT
gorilla_gorilla/1-200      AAGTCTGACAGTGGCAGGCCATAATACTACTGAATAAATTCTTACCAAACTGAAGCTTCT
macaca_fascicularis/1-200  AAGACTGATGGTGGCAGGCTGTAATACTACTGAATAAATTCTTACTGAACTGAAGCTTCT
                           *** ****  *********  ************************  *************
homo_sapiens/1-200         GGGCTCAGTTTCAATGTTTTATTTAGTATTGTTCTCACCAGCTTTCTTCTGACAGGGCTC
pan_troglodytes/1-200      GGGCTCAGTTTCAATGTTTTATTTAGTATTATTCTCGCCAGCTTTCTTCTGACAGGGCTC
gorilla_gorilla/1-200      GGGCTCAGTTTCAATGTTTTATTTAGTATTATTCTCGCCAGCTTTCTTCTGACAGGGCTC
macaca_fascicularis/1-200  GGGCTCAGTTTCAATGTTTTCTTTAGTATTATTGTCGCCAGCTTTCTTCTGACAGGGCTC
                           ******************** ********* ** ** ************************
homo_sapiens/1-200         AATGAAGATGTTAGTTAATA
pan_troglodytes/1-200      AATCAAGATGTTAGTTAATA
gorilla_gorilla/1-200      AATCAAGATGTTAGTTAATA
macaca_fascicularis/1-200  AATCAAGATGTTAGTTAATA
                           *** ****************
```

ANC1208:

- Intron 11, 289bp
- **Hg38:** chr7:147516200-147516488
- **Pan_tro_3.0:** chr7:152029381-152029669
- **gorGor4:** chr7:147120083-147120371
- **Macaca_fascicularis_5.0:** chr3:180495020-180495310

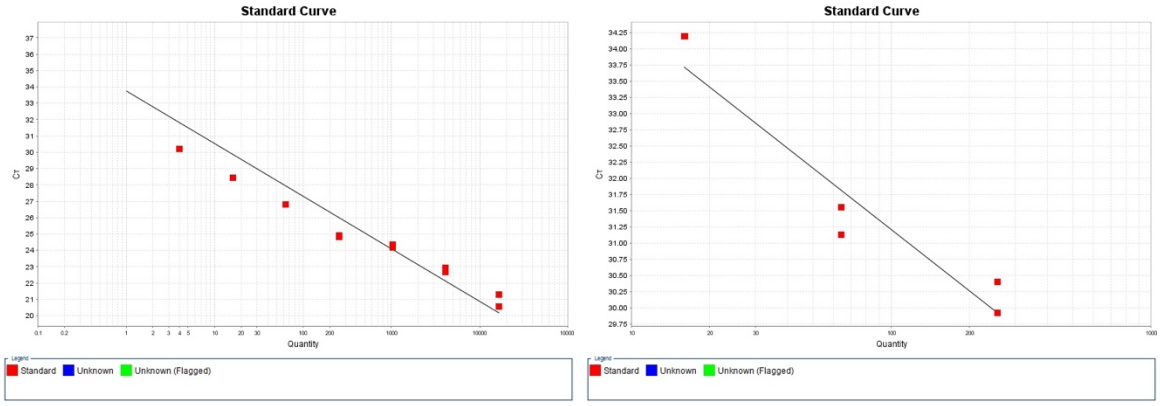
```
homo_sapiens/1-310         TTA---AAAGTTGTCTAT--TGCTAATTATATATTT-GTACTT----TT-TTTTA-CAGC
pan_troglodytes/1-310      TTA---AAAATTGTCTAT--TGCTAATTATATATTT-GTACTT----TT-TTTTA-CAGC
gorilla_gorilla/1-310      TTA---AAAATTGTCTAT--TGCTAATTATATATTT-GTACTT----TT-TTTTA-TAGC
macaca_fascicularis/1-310 ---ATAAAAATTA----T--TGCTAATTATATATTT-GTACTTGCTTTT-TTTTA-TAGC
                                *** **     *  *************** ******    ** *****  ***
homo_sapiens/1-310         TACTCATGTGTAGAACAAAAAAATA--GACAAAATTACTCA--GGCTCAATGATTCACTA
pan_troglodytes/1-310      TACTCATGTGTAGAACAAAAAAATA--GACAAAATTACTCA--GGCTCAATGATTCACTA
gorilla_gorilla/1-310      TACTCATGTGTAGAACAAAAAAATA--GACAAAATTACTCA--GGCTCAATGATTCACTA
macaca_fascicularis/1-310 TACTCATGTGTAGAACAAAAAAATA--GACAAAATTAATCATAGGCTCAATGATTCACTA
                           *************************  ********** ***  ****************
homo_sapiens/1-310         TGTACAGTTTACAAATTAATT----AGCTT-CTCTTAATTTTTTCAATTAAAAAAGTGTT
pan_troglodytes/1-310      TGTACAGTTTACAAATTAATT----AGCTT-CTCTTAATCTTTTCAATTAAAAAAGTGTT
gorilla_gorilla/1-310      TGTACAGTTTACAAATTAATT----AGCTT-CTCTTAATCTTTTCAATTAAAAAAGTGTT
macaca_fascicularis/1-310 TGTACAGTTTACAAATT----AATTAGCTT-CTCTTAATCTTTTCAATTAAAAAAGTGTT
                           *****************        ***** ******** ******************
homo_sapiens/1-310         AAAGCATTAGTTTATTGCTTATTTCCGTGTTAAAATGCTGGTTAATTGCAAAACATTATT
pan_troglodytes/1-310      AAAGCATTAGTTTATTGCTTACTTCCGTGTTAAAATGCTGGTTAATTGCAAAACATTATT
gorilla_gorilla/1-310      AAAGCATTAGTTTATTGCTTACTTCCGTGTTAAAATGCTGGTTAATTGCAAAACATTATT
macaca_fascicularis/1-310 AAAGCATTAGTTTATTGCTTACTTCCGTGTTAAAATGCTGGTTAATTGCAAAACATTATT
                           ******************** ************************************
homo_sapiens/1-310         AGGGGCCAATATTAAAATTCCACTTACAAGTGTCAGAGAAGCAAAACACATTTAAAGTTC
pan_troglodytes/1-310      AGGGGCCAATATTAAAATTCCACTTACAAGTGTCAGAGAAGCAAAACACATTTAAAGTTG
gorilla_gorilla/1-310      AGGGGCCAATATTAAAATTCCACTTACAAGTGTCAGAGAAGCAAAACACATTTAAAGTTG
```

```
macaca_fascicularis/1-310 AGGGGCCAATATTAAAATTCCACTTACAAGTGTCAGAGAAGCAAAACACATTTAAAGTTG
                          ************************************************************
homo_sapiens/1-310        ATTTGCAAAT
pan_troglodytes/1-310     ATTTGCAAAT
gorilla_gorilla/1-310     ATTTGCACAT
macaca_fascicularis/1-310 ATTTGCAAAT
                          ******* **
```

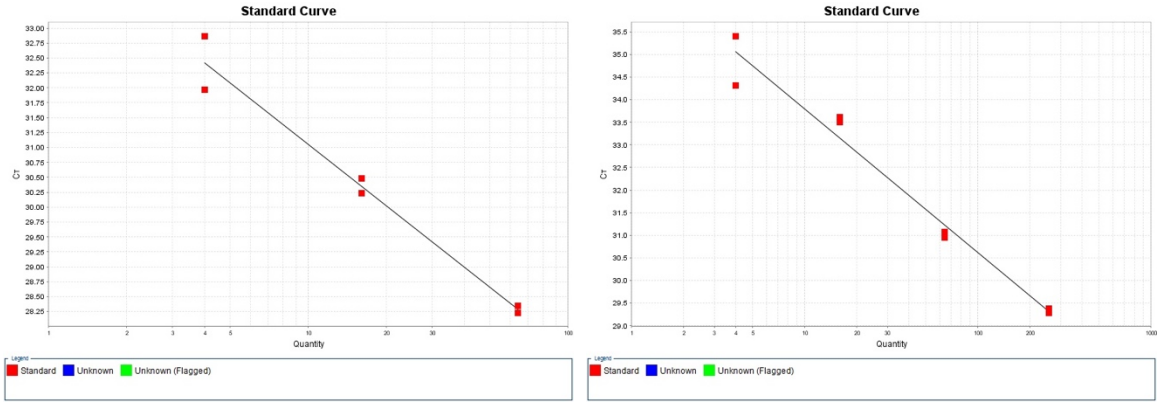
Appendix 3: *Human and macaque qRT-PCR primer efficiency curves*

Exon 8 primer pair

Human *Macaque*



Exon 18 primer pair



Appendix 4: scRNA-Seq cluster markers

p_val: unadjusted *p*-value
avg_logFC: log(WT/KO expression)
pct.1: percentage of WT cells in the cluster expressing the gene
pct.2: percentage of KO cells in the cluster expressing the gene
p_val_adj: *p*-value after correction for multiple testing

Pre-thalamus

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
0.00E+00	1.8069895	0.718	0.158	0.00E+00	0	SIX3
0.00E+00	1.4956744	0.345	0.03	0.00E+00	0	DLX6-AS1
0.00E+00	1.4779589	0.499	0.042	0.00E+00	0	DLX5
0.00E+00	1.4267599	0.605	0.199	0.00E+00	0	MEIS2
0.00E+00	1.2051834	0.401	0.026	0.00E+00	0	DLX6
0.00E+00	1.1737787	0.55	0.142	0.00E+00	0	MIR7-3HG
0.00E+00	1.162414	0.938	0.546	0.00E+00	0	STMN2
0.00E+00	1.1291214	0.318	0.018	0.00E+00	0	ISL1
0.00E+00	1.1129284	0.37	0.061	0.00E+00	0	SIX3-AS1
0.00E+00	1.0930775	0.808	0.514	0.00E+00	0	NR2F1
0.00E+00	1.0779833	0.715	0.369	0.00E+00	0	GNG3
0.00E+00	1.0674115	0.912	0.589	0.00E+00	0	STMN4
0.00E+00	1.0444821	0.344	0.033	0.00E+00	0	DLX2
0.00E+00	1.0067095	0.34	0.03	0.00E+00	0	GAD2
0.00E+00	0.9822447	0.444	0.131	0.00E+00	0	CADM2
0.00E+00	0.9481276	0.309	0.022	0.00E+00	0	DLX1
0.00E+00	0.8874315	0.876	0.593	0.00E+00	0	CD24
0.00E+00	0.704169	0.24	0.024	0.00E+00	0	SLC32A1
4.18E-276	0.6176047	0.903	0.778	1.40E-271	0	MLLT11
3.19E-272	0.8700309	0.615	0.318	1.07E-267	0	NR2F2
1.20E-237	0.5930762	0.172	0.017	4.02E-233	0	ESRRG
9.77E-231	0.9132565	0.555	0.299	3.28E-226	0	MIAT
1.56E-227	0.7785167	0.308	0.082	5.24E-223	0	MIR124-2HG
8.98E-226	0.8152002	0.292	0.075	3.01E-221	0	POU3F4
1.28E-222	0.9074994	0.351	0.115	4.30E-218	0	RBP1
2.45E-213	0.8778584	0.352	0.122	8.21E-209	0	DCBLD2
2.62E-213	0.6680125	0.247	0.053	8.79E-209	0	POU3F2
4.86E-206	1.2267356	0.622	0.47	1.63E-201	0	PAX6
5.28E-202	0.5315701	0.883	0.78	1.77E-197	0	TUBB2A
1.23E-198	0.7664466	0.268	0.07	4.13E-194	0	FIGN

8.17E-197	0.8385813	0.655	0.501	2.74E-192	0	NOVA1
1.15E-176	0.770787	0.562	0.312	3.86E-172	0	CELF4
1.26E-174	0.459737	0.127	0.012	4.22E-170	0	KCNMB2
7.02E-171	0.669413	0.734	0.603	2.35E-166	0	SYT1
1.95E-169	0.720161	0.291	0.094	6.53E-165	0	RUNX1T1
2.62E-166	0.500083	0.159	0.025	8.77E-162	0	FOXP2
2.21E-165	0.641902	0.665	0.49	7.40E-161	0	TAGLN3
5.98E-159	0.718898	0.329	0.129	2.01E-154	0	ZFHX3
4.07E-151	0.642989	0.582	0.37	1.36E-146	0	ELAVL4
1.10E-150	0.679216	0.689	0.581	3.71E-146	0	DAAM1
6.32E-149	0.635733	0.608	0.413	2.12E-144	0	DCX
3.47E-121	0.600048	0.47	0.277	1.16E-116	0	INA
4.15E-118	0.591558	0.568	0.408	1.39E-113	0	KIF5C
1.37E-111	0.494495	0.691	0.571	4.60E-107	0	CRMP1
1.06E-110	0.41649	0.135	0.028	3.54E-106	0	CDH8
6.97E-106	0.469067	0.652	0.473	2.34E-101	0	RAB3A
2.91E-104	0.517145	0.313	0.144	9.75E-100	0	LMO3
1.10E-102	0.821947	0.414	0.265	3.69E-98	0	CASP3
2.59E-98	0.540313	0.501	0.342	8.68E-94	0	GDAP1L1
1.32E-97	0.407484	0.72	0.571	4.43E-93	0	GAP43
2.07E-96	0.560559	0.359	0.193	6.95E-92	0	SCG3
3.30E-92	0.435588	0.164	0.05	1.11E-87	0	DYNC1I1
2.15E-88	0.615917	0.325	0.18	7.19E-84	0	SLF1
1.73E-87	0.523142	0.578	0.471	5.80E-83	0	NSG1
2.08E-85	0.60928	0.282	0.144	6.97E-81	0	ARX
6.29E-83	0.422676	0.152	0.047	2.11E-78	0	ZNF385D
7.75E-82	0.525748	0.384	0.232	2.60E-77	0	CELF3
6.31E-80	0.416364	0.153	0.049	2.12E-75	0	RCAN2
1.10E-76	0.313917	0.694	0.498	3.71E-72	0	RTN1
1.44E-76	0.595657	0.433	0.315	4.84E-72	0	GPC2
1.56E-74	0.534919	0.417	0.275	5.23E-70	0	ATP1A3
1.89E-74	0.677505	0.386	0.259	6.33E-70	0	TSHZ2
2.24E-68	0.613427	0.4	0.29	7.52E-64	0	MUM1
2.60E-67	0.500704	0.206	0.091	8.70E-63	0	NEFL
3.03E-65	0.524226	0.323	0.197	1.01E-60	0	PPP1R1A
7.23E-63	0.497266	0.495	0.393	2.43E-58	0	RBFOX2
7.32E-61	0.454391	0.23	0.114	2.46E-56	0	ASCL1
4.65E-60	0.58448	0.318	0.207	1.56E-55	0	XPR1
1.87E-57	0.582612	0.314	0.205	6.27E-53	0	BAZ2B

2.01E-50	0.436739	0.388	0.273	6.73E-46	0	MAPT
2.85E-18	0.299635	0.594	0.696	9.55E-14	0	AP2M1
5.95E-13	0.271735	0.562	0.672	2.00E-08	0	PAIP2
2.14E-06	0.279835	0.527	0.641	7.16E-02	0	ARF4
5.31E-06	0.253069	0.49	0.597	1.78E-01	0	PRRC2C
1.44E-05	0.258897	0.465	0.566	4.84E-01	0	SMARCB1
9.45E-05	0.274376	0.469	0.571	1.00E+00	0	SEC62

Cortical progenitors

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
0.00E+00	1.189567	0.914	0.421	0.00E+00	1	ID3
0.00E+00	1.128655	0.803	0.17	0.00E+00	1	SFRP1
0.00E+00	1.115153	0.993	0.697	0.00E+00	1	C1orf61
0.00E+00	1.05932	0.922	0.507	0.00E+00	1	ID1
0.00E+00	1.016335	0.923	0.363	0.00E+00	1	ZFP36L1
0.00E+00	0.823635	0.992	0.861	0.00E+00	1	NAP1L1
0.00E+00	0.818156	0.573	0.077	0.00E+00	1	AMBN
0.00E+00	0.807691	0.82	0.327	0.00E+00	1	HES1
0.00E+00	0.797807	0.558	0.093	0.00E+00	1	IFI44L
0.00E+00	0.795402	0.737	0.248	0.00E+00	1	DOK5
0.00E+00	0.783057	0.653	0.18	0.00E+00	1	MPPED2
0.00E+00	0.777292	0.646	0.181	0.00E+00	1	HMGA2
0.00E+00	0.770947	0.584	0.177	0.00E+00	1	PCLAF
0.00E+00	0.767658	0.918	0.399	0.00E+00	1	SOX2
0.00E+00	0.759532	0.666	0.223	0.00E+00	1	CCND2
0.00E+00	0.753716	0.999	0.863	0.00E+00	1	VIM
0.00E+00	0.728338	0.605	0.151	0.00E+00	1	SOX3
0.00E+00	0.722099	0.668	0.262	0.00E+00	1	MCM7
0.00E+00	0.712741	0.65	0.247	0.00E+00	1	FGFBP3
0.00E+00	0.710875	0.703	0.257	0.00E+00	1	EFNB1
0.00E+00	0.709381	0.932	0.51	0.00E+00	1	EMX2
0.00E+00	0.694857	0.674	0.249	0.00E+00	1	GPC1
0.00E+00	0.659562	0.985	0.878	0.00E+00	1	NPM1
0.00E+00	0.631329	0.747	0.294	0.00E+00	1	LINC01551
0.00E+00	0.62593	0.673	0.259	0.00E+00	1	VCAN
0.00E+00	0.624499	0.485	0.105	0.00E+00	1	GAS1
0.00E+00	0.608179	0.44	0.109	0.00E+00	1	GINS2
0.00E+00	0.561735	0.433	0.096	0.00E+00	1	CDT1
0.00E+00	0.559376	0.43	0.108	0.00E+00	1	COL9A3
0.00E+00	0.558956	0.981	0.658	0.00E+00	1	TTYH1
0.00E+00	0.52908	0.353	0.064	0.00E+00	1	MCM5
0.00E+00	0.490262	0.399	0.09	0.00E+00	1	LEF1
5.75E-308	0.499828	0.367	0.076	1.93E-303	1	SNCG
2.37E-296	0.61779	0.747	0.336	7.96E-292	1	PHGDH
1.05E-282	0.607799	0.615	0.233	3.53E-278	1	HEY1
2.74E-282	0.475378	0.398	0.101	9.18E-278	1	EFHD1

1.76E-269	0.56548	0.712	0.319	5.90E-265	1	FBLN1
3.37E-265	0.550665	0.521	0.184	1.13E-260	1	SOX9
1.56E-264	0.466519	0.372	0.091	5.23E-260	1	LIX1
1.06E-262	0.409044	0.319	0.065	3.56E-258	1	PRKCB
3.61E-262	0.630076	0.813	0.445	1.21E-257	1	ID2
1.41E-261	0.529699	0.539	0.191	4.72E-257	1	B3GAT2
3.32E-259	0.55911	0.762	0.398	1.11E-254	1	FBL
9.43E-259	0.483087	0.767	0.345	3.16E-254	1	LHX2
2.34E-256	0.541614	0.934	0.64	7.86E-252	1	FABP7
1.66E-255	0.433101	0.301	0.061	5.58E-251	1	MCM2
2.82E-254	0.412482	0.303	0.063	9.45E-250	1	GLI3
6.31E-252	0.594446	0.823	0.486	2.12E-247	1	GNAI2
3.08E-246	0.574679	0.525	0.189	1.03E-241	1	TYMS
1.20E-242	0.461534	0.957	0.783	4.03E-238	1	PABPC1
8.98E-240	0.441204	0.973	0.804	3.01E-235	1	EEF1D
1.74E-239	0.539749	0.875	0.571	5.83E-235	1	HSPD1
2.33E-233	0.453951	0.563	0.214	7.80E-229	1	MASP1
5.66E-230	0.446755	0.332	0.083	1.90E-225	1	MCM3
1.86E-226	0.51693	0.843	0.539	6.24E-222	1	IMPDH2
3.86E-226	0.519023	0.617	0.276	1.29E-221	1	CMBL
1.30E-224	0.549934	0.8	0.469	4.38E-220	1	GLUL
6.87E-218	0.488183	0.615	0.279	2.30E-213	1	TMEM123
3.66E-216	0.491996	0.771	0.432	1.23E-211	1	RSL1D1
4.13E-214	0.495315	0.633	0.297	1.38E-209	1	PAICS
5.21E-214	0.452702	0.953	0.757	1.75E-209	1	EEF1B2
3.34E-208	0.417287	0.371	0.112	1.12E-203	1	CDCA7L
6.79E-207	0.478576	0.606	0.269	2.28E-202	1	ID4
2.46E-202	0.421375	0.384	0.122	8.26E-198	1	CTSC
3.00E-201	0.486747	0.677	0.348	1.01E-196	1	CDK4
4.95E-195	0.514732	0.633	0.317	1.66E-190	1	RPL22L1
7.48E-193	0.451896	0.435	0.156	2.51E-188	1	CENPH
3.81E-191	0.491631	0.679	0.346	1.28E-186	1	PSAT1
7.84E-191	0.463416	0.76	0.435	2.63E-186	1	CCNB1IP1
2.48E-189	0.381503	0.346	0.107	8.33E-185	1	ADA
4.54E-189	0.478044	0.8	0.487	1.52E-184	1	TMEM161B-AS1
4.02E-188	0.404283	0.527	0.22	1.35E-183	1	CCND1
1.82E-187	0.430248	0.436	0.163	6.11E-183	1	C19orf48
3.40E-186	0.435969	0.923	0.706	1.14E-181	1	SERBP1
1.01E-179	0.349793	0.967	0.854	3.40E-175	1	SLC25A3

1.41E-177	0.427474	0.386	0.138	4.74E-173	1	EMP2
3.14E-177	0.426209	0.45	0.18	1.05E-172	1	PGM1
1.54E-173	0.364379	0.28	0.076	5.18E-169	1	BOC
1.01E-171	0.328409	0.281	0.076	3.40E-167	1	MTHFD1
1.48E-169	0.365379	0.344	0.113	4.96E-165	1	TP53
1.53E-169	0.452961	0.752	0.438	5.13E-165	1	SNRPB
9.84E-169	0.280651	0.219	0.047	3.30E-164	1	MCM4
1.14E-168	0.330151	0.232	0.054	3.82E-164	1	TCIM
1.63E-168	0.433082	0.527	0.235	5.46E-164	1	SYNE2
8.64E-163	0.412281	0.533	0.248	2.90E-158	1	AHCY
1.34E-161	0.343474	0.979	0.853	4.49E-157	1	PKM
1.27E-160	0.510432	0.497	0.215	4.27E-156	1	AC007952.4
3.14E-154	0.458708	0.703	0.402	1.05E-149	1	NASP
4.21E-154	0.288238	0.233	0.057	1.41E-149	1	LRRC3B
5.03E-153	0.399914	0.499	0.226	1.69E-148	1	IVNS1ABP
5.75E-153	0.385362	0.898	0.683	1.93E-148	1	CNBP
5.40E-151	0.412791	0.708	0.404	1.81E-146	1	SSRP1
3.17E-150	0.369666	0.4	0.158	1.06E-145	1	GPC4

Cortical hem

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
0.00E+00	1.75421	0.856	0.124	0.00E+00	2	RSPO3
0.00E+00	1.652063	0.333	0.048	0.00E+00	2	SPARCL1
0.00E+00	1.360051	0.985	0.576	0.00E+00	2	CLU
0.00E+00	1.295217	0.928	0.262	0.00E+00	2	WLS
0.00E+00	1.288214	0.833	0.127	0.00E+00	2	RSPO2
0.00E+00	1.262757	0.831	0.402	0.00E+00	2	SAT1
0.00E+00	1.222843	0.701	0.17	0.00E+00	2	NMU
0.00E+00	1.183365	0.715	0.119	0.00E+00	2	NTRK2
0.00E+00	1.180076	0.741	0.114	0.00E+00	2	TPBG
0.00E+00	1.167949	0.732	0.202	0.00E+00	2	IFITM3
0.00E+00	1.167166	0.866	0.275	0.00E+00	2	ANXA2
0.00E+00	1.159586	0.903	0.332	0.00E+00	2	SPARC
0.00E+00	1.139864	0.306	0.026	0.00E+00	2	LPL
0.00E+00	1.08441	0.82	0.277	0.00E+00	2	PLTP
0.00E+00	1.039479	0.939	0.437	0.00E+00	2	CD99
0.00E+00	1.030437	0.596	0.118	0.00E+00	2	S100A10
0.00E+00	1.020011	0.986	0.671	0.00E+00	2	TTYH1
0.00E+00	0.994827	0.849	0.309	0.00E+00	2	RHOC
0.00E+00	0.985759	0.774	0.197	0.00E+00	2	SDC2
0.00E+00	0.982217	0.675	0.124	0.00E+00	2	MGST1
0.00E+00	0.960747	0.806	0.289	0.00E+00	2	IGFBP5
0.00E+00	0.934389	0.48	0.096	0.00E+00	2	BAMBI
0.00E+00	0.909842	0.66	0.199	0.00E+00	2	LYPD1
0.00E+00	0.898808	0.947	0.511	0.00E+00	2	CRABP2
0.00E+00	0.898187	0.396	0.046	0.00E+00	2	S100A11
0.00E+00	0.897712	0.964	0.662	0.00E+00	2	MDK
0.00E+00	0.881752	0.764	0.252	0.00E+00	2	CDO1
0.00E+00	0.835456	0.979	0.753	0.00E+00	2	ITM2B
0.00E+00	0.831744	0.564	0.068	0.00E+00	2	RSPO1
0.00E+00	0.830135	0.77	0.281	0.00E+00	2	CYSTM1
0.00E+00	0.825304	0.64	0.163	0.00E+00	2	PON2
0.00E+00	0.819311	0.685	0.256	0.00E+00	2	TAGLN2
0.00E+00	0.803581	0.905	0.454	0.00E+00	2	NPC2
0.00E+00	0.796971	0.943	0.627	0.00E+00	2	CALR
0.00E+00	0.785914	0.933	0.517	0.00E+00	2	MEST
0.00E+00	0.760022	0.558	0.148	0.00E+00	2	PLS3

0.00E+00	0.753767	0.677	0.214	0.00E+00	2	SERPINH1
0.00E+00	0.747877	0.996	0.892	0.00E+00	2	SERF2
0.00E+00	0.744283	0.531	0.093	0.00E+00	2	KRT8
0.00E+00	0.738496	0.523	0.138	0.00E+00	2	NEAT1
0.00E+00	0.737972	0.426	0.071	0.00E+00	2	FAM107A
0.00E+00	0.721076	0.426	0.049	0.00E+00	2	CFAP126
0.00E+00	0.708605	0.39	0.046	0.00E+00	2	EDNRB
0.00E+00	0.693529	0.6	0.183	0.00E+00	2	GSN
0.00E+00	0.682153	0.949	0.575	0.00E+00	2	HSPB1
0.00E+00	0.679466	0.557	0.158	0.00E+00	2	TGIF1
0.00E+00	0.675416	0.993	0.849	0.00E+00	2	GSTP1
0.00E+00	0.668768	0.56	0.134	0.00E+00	2	ZNF503
0.00E+00	0.657861	0.424	0.059	0.00E+00	2	C5orf49
0.00E+00	0.648765	0.51	0.092	0.00E+00	2	SLIT2
0.00E+00	0.645983	0.391	0.053	0.00E+00	2	WNT2B
0.00E+00	0.641759	0.481	0.097	0.00E+00	2	ADGRV1
0.00E+00	0.629193	0.563	0.155	0.00E+00	2	EMC2
0.00E+00	0.627103	0.436	0.061	0.00E+00	2	PIFO
0.00E+00	0.622072	0.406	0.054	0.00E+00	2	IFITM2
0.00E+00	0.607067	0.545	0.152	0.00E+00	2	GULP1
0.00E+00	0.587604	0.381	0.041	0.00E+00	2	GDPD2
0.00E+00	0.583383	0.378	0.047	0.00E+00	2	LGALS3
0.00E+00	0.571461	0.55	0.142	0.00E+00	2	MIR99AHG
0.00E+00	0.566244	0.508	0.12	0.00E+00	2	MLLT1
0.00E+00	0.557046	0.424	0.06	0.00E+00	2	HTRA1
0.00E+00	0.546879	0.352	0.038	0.00E+00	2	ATP1A2
0.00E+00	0.543916	0.314	0.036	0.00E+00	2	S100B
0.00E+00	0.528741	0.495	0.119	0.00E+00	2	LHFPL6
0.00E+00	0.521816	0.388	0.072	0.00E+00	2	PRTG
0.00E+00	0.51807	0.298	0.037	0.00E+00	2	CD44
0.00E+00	0.500824	0.352	0.06	0.00E+00	2	CD9
0.00E+00	0.46443	0.262	0.029	0.00E+00	2	ANXA1
0.00E+00	0.459557	0.301	0.031	0.00E+00	2	GCNT1
0.00E+00	0.457886	0.34	0.052	0.00E+00	2	WNT5A
0.00E+00	0.448288	0.312	0.04	0.00E+00	2	SULF1
5.85E-308	0.628571	0.613	0.196	1.96E-303	2	KLHDC8B
5.39E-303	0.556223	0.432	0.096	1.81E-298	2	KRT18
2.58E-300	0.435366	0.217	0.018	8.66E-296	2	CCDC80
3.08E-300	0.462798	0.36	0.064	1.03E-295	2	PERP

1.49E-297	0.451136	0.325	0.053	5.01E-293	2	PLA2G16
3.14E-297	0.696678	0.792	0.366	1.05E-292	2	CETN2
9.38E-297	0.705643	0.857	0.468	3.15E-292	2	CST3
2.52E-295	0.502263	0.412	0.087	8.44E-291	2	OTX2
3.20E-293	0.676828	0.882	0.504	1.07E-288	2	IFI27L2
1.02E-292	0.633405	0.58	0.186	3.42E-288	2	PLEKHA5
1.20E-292	0.376889	0.27	0.035	4.04E-288	2	PPIC
3.12E-292	0.488672	0.423	0.096	1.05E-287	2	TNFRSF1A
1.30E-291	0.919465	0.514	0.154	4.35E-287	2	LIPA
3.65E-291	0.40762	0.277	0.037	1.22E-286	2	RSPH1
4.08E-288	0.564154	0.672	0.223	1.37E-283	2	TSHZ2
4.55E-288	0.618056	0.603	0.206	1.53E-283	2	OBSL1
3.10E-286	0.622454	0.625	0.22	1.04E-281	2	GRN
2.32E-284	0.405708	0.292	0.042	7.77E-280	2	ADCY2
5.52E-281	0.604121	0.655	0.241	1.85E-276	2	CTSD
2.95E-280	0.624619	0.655	0.248	9.88E-276	2	MLF1
3.26E-279	0.575772	0.531	0.16	1.09E-274	2	PLPP3
9.07E-279	0.663219	0.727	0.31	3.04E-274	2	TIMP1
1.74E-278	0.559308	0.531	0.16	5.82E-274	2	CIB1
2.82E-277	0.639356	0.898	0.425	9.45E-273	2	SOX2
1.20E-276	0.51549	0.469	0.125	4.01E-272	2	HSD17B8
4.54E-276	0.49764	0.443	0.111	1.52E-271	2	ANXA6
7.27E-275	0.525392	0.3	0.047	2.44E-270	2	CA4
2.44E-273	0.73774	0.431	0.103	8.20E-269	2	CLDN5
8.92E-273	0.440516	0.22	0.023	2.99E-268	2	IFIT1
6.58E-272	0.426933	0.347	0.067	2.21E-267	2	LRP10
4.18E-271	0.404988	0.308	0.052	1.40E-266	2	MAF
6.36E-269	0.395319	0.214	0.021	2.13E-264	2	LUM
5.75E-265	0.441114	0.307	0.053	1.93E-260	2	GJA1
3.52E-264	1.485706	0.555	0.203	1.18E-259	2	TRH
3.50E-262	0.699499	0.397	0.093	1.17E-257	2	APOE
1.01E-258	0.605427	0.585	0.204	3.37E-254	2	CHPF
6.05E-256	0.578204	0.809	0.35	2.03E-251	2	HES1
1.74E-255	0.546158	0.569	0.193	5.84E-251	2	SLC44A2
4.79E-254	0.485698	0.337	0.069	1.61E-249	2	GNG11
9.69E-254	0.895664	0.585	0.226	3.25E-249	2	MASP1
3.02E-252	0.596185	0.764	0.354	1.01E-247	2	NENF
1.65E-251	0.442754	0.292	0.049	5.54E-247	2	LGI1
3.90E-251	0.428076	0.343	0.07	1.31E-246	2	KIF9

1.70E-250	0.611768	0.409	0.105	5.71E-246	2	OLFM3
1.21E-249	1.197935	0.762	0.363	4.04E-245	2	PTN
7.51E-246	0.557219	0.425	0.112	2.52E-241	2	SPRY1
2.19E-245	0.607034	0.602	0.224	7.35E-241	2	CRB2
3.88E-245	0.361428	0.268	0.042	1.30E-240	2	DMKN
6.98E-240	0.452851	0.346	0.076	2.34E-235	2	SELENOP
4.40E-238	0.458896	0.425	0.113	1.47E-233	2	ADD3
4.55E-238	0.306689	0.227	0.029	1.53E-233	2	LINGO2
1.47E-236	0.454292	0.285	0.05	4.93E-232	2	TSTD1
3.04E-236	0.510349	0.349	0.077	1.02E-231	2	MSX1

Excitatory neurons

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
0.00E+00	2.240489	0.986	0.214	0.00E+00	3	NEUROD6
0.00E+00	1.738986	0.894	0.097	0.00E+00	3	NEUROD2
0.00E+00	1.727241	0.68	0.088	0.00E+00	3	CALB2
0.00E+00	1.703793	0.955	0.34	0.00E+00	3	NFIB
0.00E+00	1.49482	0.763	0.057	0.00E+00	3	THSD7A
0.00E+00	1.373713	0.91	0.34	0.00E+00	3	NFIA
0.00E+00	1.340726	0.758	0.094	0.00E+00	3	SSTR2
0.00E+00	1.336807	0.822	0.151	0.00E+00	3	CAMKV
0.00E+00	1.151834	0.88	0.365	0.00E+00	3	BCL11A
0.00E+00	1.148442	0.988	0.469	0.00E+00	3	RTN1
0.00E+00	1.02461	0.723	0.127	0.00E+00	3	TBR1
0.00E+00	1.022982	0.844	0.291	0.00E+00	3	OLFM1
0.00E+00	0.956333	0.43	0.032	0.00E+00	3	BHLHE22
0.00E+00	0.952432	0.547	0.064	0.00E+00	3	ZBTB18
0.00E+00	0.952423	0.93	0.545	0.00E+00	3	CNTNAP2
0.00E+00	0.93484	0.964	0.545	0.00E+00	3	GAP43
0.00E+00	0.926312	0.762	0.309	0.00E+00	3	ABRACL
0.00E+00	0.899687	0.518	0.115	0.00E+00	3	RASL11B
0.00E+00	0.898036	0.804	0.322	0.00E+00	3	THRA
0.00E+00	0.896974	0.721	0.258	0.00E+00	3	NRN1
0.00E+00	0.896668	0.473	0.083	0.00E+00	3	BCHE
0.00E+00	0.865861	0.69	0.284	0.00E+00	3	LMO4
0.00E+00	0.862614	0.801	0.343	0.00E+00	3	CELF2
0.00E+00	0.851133	0.713	0.245	0.00E+00	3	SNCA
0.00E+00	0.838058	0.842	0.288	0.00E+00	3	CELF4
0.00E+00	0.815423	0.482	0.041	0.00E+00	3	SNCB
0.00E+00	0.796547	0.906	0.449	0.00E+00	3	RAB3A
0.00E+00	0.787666	0.56	0.113	0.00E+00	3	ZEB2
0.00E+00	0.772892	0.389	0.046	0.00E+00	3	GUCY1A1
0.00E+00	0.771066	0.98	0.773	0.00E+00	3	TUBB2A
0.00E+00	0.768481	0.442	0.065	0.00E+00	3	NHLH2
0.00E+00	0.767333	0.567	0.15	0.00E+00	3	CAP2
0.00E+00	0.763108	0.925	0.616	0.00E+00	3	FEZ1
0.00E+00	0.734813	0.918	0.546	0.00E+00	3	CRMP1
0.00E+00	0.728466	0.583	0.125	0.00E+00	3	GRIA2
0.00E+00	0.720636	0.558	0.167	0.00E+00	3	NSMF

0.00E+00	0.718539	0.29	0.032	0.00E+00	3	DIRAS3
0.00E+00	0.718373	0.619	0.184	0.00E+00	3	IGSF21
0.00E+00	0.717278	0.441	0.038	0.00E+00	3	SLC17A7
0.00E+00	0.705126	0.897	0.52	0.00E+00	3	TMEM59L
0.00E+00	0.704406	0.966	0.692	0.00E+00	3	GPM6A
0.00E+00	0.68247	0.961	0.775	0.00E+00	3	CALM1
0.00E+00	0.680929	0.986	0.869	0.00E+00	3	BASP1
0.00E+00	0.677978	0.637	0.175	0.00E+00	3	TTC9B
0.00E+00	0.667823	0.587	0.176	0.00E+00	3	JPH4
0.00E+00	0.666469	0.786	0.28	0.00E+00	3	NSG2
0.00E+00	0.652083	0.411	0.068	0.00E+00	3	NEUROD1
0.00E+00	0.619811	0.575	0.152	0.00E+00	3	NRXN1
0.00E+00	0.558406	0.298	0.032	0.00E+00	3	ARPP21
0.00E+00	0.551673	0.403	0.072	0.00E+00	3	NOL4
0.00E+00	0.55145	0.376	0.066	0.00E+00	3	PHYHIP
0.00E+00	0.541113	0.439	0.088	0.00E+00	3	FXYD7
0.00E+00	0.530444	0.298	0.022	0.00E+00	3	SEZ6
0.00E+00	0.521874	0.303	0.033	0.00E+00	3	CORO2B
0.00E+00	0.515765	0.352	0.057	0.00E+00	3	B3GAT1
0.00E+00	0.507795	0.331	0.047	0.00E+00	3	SLIT1
0.00E+00	0.467408	0.302	0.037	0.00E+00	3	SH3GL2
0.00E+00	0.458714	0.229	0.004	0.00E+00	3	SLA
0.00E+00	0.448393	0.266	0.016	0.00E+00	3	MGLL
0.00E+00	0.444279	0.238	0.01	0.00E+00	3	AC004158.1
1.20E-307	0.637286	0.704	0.235	4.01E-303	3	MAPT
5.65E-303	0.538835	0.267	0.031	1.89E-298	3	SYNPR
1.65E-302	0.867236	0.664	0.275	5.54E-298	3	FEZF2
1.91E-299	0.642636	0.407	0.083	6.40E-295	3	NXPH4
4.22E-299	0.672272	0.936	0.713	1.41E-294	3	DSTN
1.13E-297	0.707818	0.757	0.33	3.80E-293	3	ATP6V1G2
4.01E-296	0.548243	0.387	0.076	1.35E-291	3	BCL11B
9.73E-296	0.498227	0.32	0.049	3.26E-291	3	RAB26
2.33E-295	0.620564	0.701	0.243	7.81E-291	3	ATP1A3
6.85E-293	0.596292	0.418	0.092	2.30E-288	3	FAM49A
9.77E-291	0.715925	0.597	0.206	3.28E-286	3	TCEAL2
4.50E-286	0.632198	0.986	0.565	1.51E-281	3	STMN2
4.45E-285	0.614256	0.529	0.145	1.49E-280	3	RGS17
1.65E-284	0.547483	0.993	0.87	5.53E-280	3	JPT1
1.69E-284	0.664969	0.674	0.254	5.66E-280	3	STXBP1

8.55E-283	0.468679	0.259	0.033	2.87E-278	3	PPP1R1B
3.03E-282	0.5925	0.541	0.156	1.01E-277	3	GNAO1
1.13E-280	1.017205	0.76	0.445	3.80E-276	3	CSRP2
2.97E-278	0.701259	0.864	0.545	9.97E-274	3	YWHAB
3.73E-278	0.632999	0.526	0.149	1.25E-273	3	ARC
4.45E-278	0.58505	0.527	0.151	1.49E-273	3	TUBB4A
6.64E-277	0.620037	0.71	0.271	2.23E-272	3	CDKN2D
4.22E-276	0.700004	0.807	0.365	1.41E-271	3	GADD45G
3.93E-268	0.637372	0.901	0.588	1.32E-263	3	YWHAH
1.95E-267	0.702026	0.712	0.313	6.54E-263	3	NELL2
3.31E-264	0.633687	0.424	0.104	1.11E-259	3	HPCA
7.50E-263	0.398535	0.282	0.041	2.52E-258	3	PHACTR3
1.51E-261	0.458944	0.321	0.056	5.06E-257	3	SLC1A2
2.03E-260	0.418134	0.298	0.047	6.82E-256	3	GABRG2
8.60E-259	0.604142	0.788	0.358	2.88E-254	3	RBFOX2
3.82E-258	0.458343	0.33	0.059	1.28E-253	3	MYT1L
3.99E-257	0.661568	0.555	0.182	1.34E-252	3	NEUROG2
4.42E-257	0.498267	0.366	0.076	1.48E-252	3	PCSK2

Pre-thalamic Eminence

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
0.00E+00	0.947274	0.298	0.024	0.00E+00	4	MAB21L1
7.49E-293	0.887304	0.899	0.49	2.51E-288	4	RTN1
5.15E-286	0.775759	0.942	0.793	1.73E-281	4	VAMP2
3.47E-278	0.780911	0.342	0.059	1.16E-273	4	NTM
9.86E-270	0.785832	0.967	0.576	3.31E-265	4	STMN2
2.83E-258	0.862757	0.706	0.3	9.49E-254	4	NSG2
9.26E-253	1.437302	0.383	0.091	3.11E-248	4	FEZF1
4.41E-242	1.255933	0.397	0.099	1.48E-237	4	LHX5-AS1
1.53E-239	0.446291	0.158	0.01	5.13E-235	4	SAMD3
2.73E-231	0.660958	0.96	0.782	9.15E-227	4	MLLT11
4.82E-219	0.628844	0.219	0.028	1.62E-214	4	LHX5
3.67E-215	0.472042	0.155	0.011	1.23E-210	4	GJD2
6.70E-206	0.72288	0.907	0.617	2.25E-201	4	STMN4
2.55E-199	0.751277	0.831	0.602	8.56E-195	4	YWHAH
1.48E-197	0.696876	0.805	0.47	4.97E-193	4	RAB3A
1.75E-192	0.811856	0.484	0.179	5.86E-188	4	ENC1
4.12E-185	0.661733	0.852	0.681	1.38E-180	4	NREP
2.35E-181	0.739883	0.434	0.149	7.88E-177	4	CAMK2N1
1.32E-174	0.552715	0.945	0.808	4.42E-170	4	PCSK1N
1.41E-174	0.735935	0.608	0.292	4.73E-170	4	CDKN2D
4.45E-170	0.564505	0.941	0.782	1.49E-165	4	TUBB2A
2.36E-168	0.61932	0.859	0.567	7.91E-164	4	GAP43
3.83E-166	0.743396	0.737	0.504	1.28E-161	4	BLCAP
1.60E-163	0.649315	0.729	0.415	5.36E-159	4	DCX
3.66E-156	0.7549	0.765	0.538	1.23E-151	4	JUN
4.24E-154	0.671327	0.765	0.493	1.42E-149	4	TAGLN3
2.49E-153	0.553893	0.88	0.617	8.35E-149	4	CD24
4.57E-146	0.535073	0.899	0.773	1.53E-141	4	BEX1
1.05E-142	0.856804	0.426	0.167	3.51E-138	4	PCDH9
1.34E-142	0.74784	0.576	0.285	4.50E-138	4	NRN1
6.97E-138	0.578669	0.818	0.636	2.34E-133	4	APLP1
6.65E-135	0.606668	0.571	0.281	2.23E-130	4	INA
1.04E-130	0.637645	0.407	0.156	3.47E-126	4	GRIA2
4.05E-125	0.608104	0.792	0.664	1.36E-120	4	ETFB
5.97E-125	0.746485	0.543	0.298	2.00E-120	4	JUNB
4.53E-124	0.593945	0.54	0.265	1.52E-119	4	MAPT

2.43E-121	0.521398	0.784	0.57	8.15E-117	4	CRMP1
1.69E-120	0.499087	0.227	0.057	5.68E-116	4	SLC17A6
8.46E-120	0.617714	0.555	0.3	2.84E-115	4	TMEM35A
3.49E-115	0.512488	0.676	0.404	1.17E-110	4	GNG3
4.71E-114	0.572068	0.737	0.547	1.58E-109	4	TMEM59L
6.96E-113	0.469711	0.867	0.71	2.33E-108	4	GPM6A
8.42E-111	0.663787	0.395	0.169	2.83E-106	4	RGS17
7.15E-110	0.553339	0.633	0.385	2.40E-105	4	RBFOX2
1.93E-107	0.732527	0.519	0.293	6.49E-103	4	FOS
5.65E-107	0.52506	0.458	0.206	1.90E-102	4	TTC9B
4.81E-105	0.624858	0.37	0.154	1.61E-100	4	SCG2
1.45E-104	0.55577	0.441	0.202	4.86E-100	4	SNAP25
4.89E-103	0.519677	0.275	0.09	1.64E-98	4	SYT4
4.02E-102	0.502296	0.297	0.105	1.35E-97	4	SYT5
6.56E-102	0.703526	0.413	0.198	2.20E-97	4	NRP2
1.06E-97	0.608028	0.637	0.454	3.56E-93	4	DNER
4.20E-97	0.537583	0.637	0.414	1.41E-92	4	KIF5C
4.20E-94	0.489924	0.676	0.463	1.41E-89	4	SEZ6L2
1.02E-91	1.044465	0.348	0.152	3.42E-87	4	NEFM
7.06E-89	0.557934	0.327	0.136	2.37E-84	4	ELAVL2
1.27E-88	0.494035	0.668	0.47	4.26E-84	4	NSG1
3.00E-87	0.492442	0.632	0.44	1.01E-82	4	KLC1
2.79E-85	0.471402	0.731	0.587	9.35E-81	4	BEX2
1.36E-83	0.435863	0.265	0.097	4.55E-79	4	ZCCHC12
2.56E-82	0.543106	0.679	0.528	8.59E-78	4	IER2
4.45E-82	0.520745	0.378	0.186	1.49E-77	4	PAK3
3.79E-80	0.461423	0.7	0.553	1.27E-75	4	CXADR
5.31E-79	0.549853	0.382	0.18	1.78E-74	4	TBR1
4.51E-78	0.471171	0.716	0.587	1.51E-73	4	DAAM1
3.27E-76	0.503033	0.554	0.363	1.10E-71	4	ATP6V1G2
7.44E-76	0.430488	0.501	0.277	2.50E-71	4	ATP1A3
4.20E-75	0.483646	0.567	0.365	1.41E-70	4	ELAVL3
1.67E-74	0.422445	0.555	0.337	5.61E-70	4	OLFM1
2.01E-74	0.48913	0.409	0.216	6.76E-70	4	CELF5
7.94E-74	0.444568	0.294	0.123	2.66E-69	4	POU2F2
3.23E-73	0.415084	0.717	0.6	1.08E-68	4	ATP6V0B
1.51E-71	0.481336	0.52	0.333	5.05E-67	4	NCAM1
1.67E-71	0.392381	0.218	0.078	5.60E-67	4	SPOCK2
1.05E-68	0.472318	0.647	0.526	3.53E-64	4	MAP1LC3A

1.34E-68	0.338984	0.144	0.037	4.48E-64	4	VGF
1.35E-67	0.606667	0.636	0.516	4.54E-63	4	NOVA1
3.90E-67	0.43862	0.246	0.098	1.31E-62	4	NOL4
1.31E-66	0.441552	0.321	0.151	4.38E-62	4	CCDC184
1.59E-66	0.427808	0.551	0.35	5.32E-62	4	GDAP1L1
5.11E-66	0.414435	0.302	0.137	1.71E-61	4	ANK3
5.70E-66	0.49362	0.42	0.24	1.91E-61	4	PKIA
4.25E-65	0.465921	0.644	0.52	1.43E-60	4	MAP2
6.88E-64	0.514718	0.242	0.096	2.31E-59	4	NEFL
2.08E-63	0.458754	0.239	0.097	6.99E-59	4	PCSK2
3.46E-63	0.743651	0.236	0.099	1.16E-58	4	PBX3
5.79E-63	0.445776	0.43	0.251	1.94E-58	4	PGM2L1
1.24E-62	0.316183	0.152	0.045	4.14E-58	4	FAM155A
2.36E-60	0.367494	0.344	0.178	7.91E-56	4	SYP
3.90E-60	0.371634	0.175	0.059	1.31E-55	4	GABRA2
9.49E-60	0.355169	0.161	0.051	3.18E-55	4	TMEFF2
1.57E-59	0.390712	0.278	0.126	5.28E-55	4	CHGB
2.58E-59	0.435029	0.423	0.24	8.65E-55	4	CELF3

Choroid Plexus

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
0.00E+00	3.905175	1	0.694	0.00E+00	5	TTR
0.00E+00	1.448323	0.713	0.1	0.00E+00	5	CLDN5
0.00E+00	1.361477	0.681	0.132	0.00E+00	5	NMB
0.00E+00	1.191017	0.896	0.164	0.00E+00	5	RSPO2
0.00E+00	1.13112	0.541	0.041	0.00E+00	5	TRPM3
0.00E+00	1.116456	0.546	0.079	0.00E+00	5	PCP4
0.00E+00	1.101059	0.992	0.765	0.00E+00	5	ITM2B
0.00E+00	1.09716	0.988	0.679	0.00E+00	5	MDK
0.00E+00	1.092515	0.779	0.155	0.00E+00	5	PLS3
0.00E+00	1.045614	0.974	0.539	0.00E+00	5	MEST
0.00E+00	1.031294	0.891	0.314	0.00E+00	5	IGFBP5
0.00E+00	0.962177	0.919	0.365	0.00E+00	5	SPARC
0.00E+00	0.95452	0.481	0.017	0.00E+00	5	HTR2C
0.00E+00	0.928674	0.334	0.028	0.00E+00	5	CXCL14
0.00E+00	0.911009	0.829	0.278	0.00E+00	5	CDO1
0.00E+00	0.899114	0.895	0.304	0.00E+00	5	WLS
0.00E+00	0.883049	0.374	0.024	0.00E+00	5	PGM5P4-AS1
0.00E+00	0.877707	0.265	0.019	0.00E+00	5	CCK
0.00E+00	0.83682	0.703	0.155	0.00E+00	5	TPBG
0.00E+00	0.80046	0.581	0.075	0.00E+00	5	MSX1
0.00E+00	0.790728	0.657	0.166	0.00E+00	5	GULP1
0.00E+00	0.758823	0.631	0.111	0.00E+00	5	KRT8
0.00E+00	0.745208	0.572	0.08	0.00E+00	5	FGFR2
0.00E+00	0.727148	0.391	0.038	0.00E+00	5	AL596223.2
0.00E+00	0.708939	0.539	0.091	0.00E+00	5	SERINC5
0.00E+00	0.688166	0.546	0.096	0.00E+00	5	OTX2
0.00E+00	0.663953	0.557	0.098	0.00E+00	5	RSPO1
0.00E+00	0.643529	0.46	0.046	0.00E+00	5	RBM47
0.00E+00	0.561239	0.429	0.062	0.00E+00	5	WNT5A
0.00E+00	0.545122	0.389	0.043	0.00E+00	5	FOLR1
1.93E-307	0.737462	0.625	0.147	6.46E-303	5	ZFYVE16
3.31E-307	0.591351	0.483	0.083	1.11E-302	5	PRTG
1.19E-305	0.856017	0.845	0.305	3.99E-301	5	CYSTM1
2.53E-299	0.757234	0.735	0.199	8.47E-295	5	NMU
4.64E-296	0.676847	0.55	0.114	1.56E-291	5	SLIT2
5.04E-295	0.429878	0.312	0.033	1.69E-290	5	FAM81B

1.34E-290	0.884705	0.94	0.479	4.49E-286	5	NPC2
8.61E-284	0.626308	0.386	0.055	2.89E-279	5	CA4
1.17E-280	0.513429	0.462	0.081	3.91E-276	5	PIFO
3.98E-277	0.834404	0.834	0.309	1.34E-272	5	PLTP
4.70E-273	0.632177	0.518	0.109	1.58E-268	5	KRT18
1.36E-271	0.727012	0.757	0.233	4.55E-267	5	SDC2
8.92E-270	0.712376	0.634	0.16	2.99E-265	5	MGST1
8.05E-269	0.593521	0.354	0.049	2.70E-264	5	SNTG1
1.17E-266	0.787515	0.717	0.222	3.91E-262	5	LYPD1
3.04E-265	0.290393	0.161	0.006	1.02E-260	5	PGM5P3-AS1
1.13E-260	0.603888	0.624	0.155	3.81E-256	5	ZNF503
3.50E-258	0.443145	0.347	0.049	1.17E-253	5	DMKN
9.32E-258	0.537543	0.431	0.078	3.12E-253	5	C3orf58
2.04E-255	0.712231	0.698	0.214	6.85E-251	5	KLHDC8B
9.93E-250	0.825172	0.973	0.601	3.33E-245	5	CLU
2.53E-245	0.555381	0.53	0.125	8.50E-241	5	PDLIM5
4.29E-243	0.491344	0.365	0.058	1.44E-238	5	LGI1
6.92E-243	0.675105	0.441	0.088	2.32E-238	5	CA2
2.27E-242	0.396085	0.288	0.035	7.62E-238	5	LINC00982
5.17E-240	0.813522	0.951	0.537	1.73E-235	5	CRABP2
4.40E-239	0.661725	0.994	0.858	1.48E-234	5	GSTP1
2.42E-238	0.404286	0.249	0.026	8.10E-234	5	CLDN3
7.47E-236	0.456867	0.334	0.05	2.51E-231	5	ARHGAP29
6.41E-233	0.806415	0.93	0.567	2.15E-228	5	MRPS6

Immature excitatory

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
7.09E-59	1.13989	0.72	0.087	2.38E-54	1	SOX2
3.40E-35	0.886393	0.591	0.1	1.14E-30	1	ZFP36L1
9.82E-34	1.800237	0.935	0.706	3.29E-29	1	VIM
1.95E-29	1.043108	0.624	0.139	6.54E-25	1	RGS16
1.43E-28	0.309681	0.204	0.004	4.79E-24	1	RGR
7.71E-27	0.749517	0.419	0.063	2.59E-22	1	NEUROG1
8.04E-27	0.404428	1	1	2.69E-22	1	RPS18
6.67E-25	0.364341	1	1	2.24E-20	1	RPL41
1.47E-24	0.699434	1	0.896	4.94E-20	1	NPM1
5.14E-24	0.505275	1	0.993	1.72E-19	1	RPL12
1.22E-23	1.64754	0.935	0.596	4.11E-19	1	HES6
2.65E-22	0.356827	1	0.999	8.87E-18	1	RPS15
4.68E-22	0.307399	0.301	0.034	1.57E-17	1	PON2
1.72E-21	0.544117	0.452	0.091	5.78E-17	1	MASP1
8.56E-21	0.355548	1	0.999	2.87E-16	1	RPL7A
9.22E-21	0.704314	1	0.77	3.09E-16	1	RPS27L
1.77E-20	0.348443	1	1	5.95E-16	1	RPS19
2.69E-20	0.360337	0.323	0.045	9.01E-16	1	CRB2
3.05E-20	0.328637	1	1	1.02E-15	1	RPL8
4.18E-20	0.477319	0.333	0.051	1.40E-15	1	MYCL
4.29E-20	0.389567	0.989	0.996	1.44E-15	1	RPSA
9.74E-20	0.363899	0.29	0.037	3.26E-15	1	AMBN
1.66E-19	0.932716	0.71	0.261	5.56E-15	1	CDKN1A
1.73E-19	0.341007	1	1	5.79E-15	1	RPS2
2.48E-19	0.263739	1	1	8.31E-15	1	RPL13A
2.75E-19	0.32898	1	1	9.22E-15	1	RPS14
2.76E-19	0.345124	1	0.999	9.26E-15	1	RPL5
3.71E-19	0.338523	1	1	1.24E-14	1	RPLP1
5.09E-19	0.901034	0.581	0.185	1.71E-14	1	DLL1
5.51E-19	0.394457	0.333	0.053	1.85E-14	1	SERPINH1
5.61E-19	0.334077	0.29	0.039	1.88E-14	1	ADA
1.48E-18	0.291084	1	1	4.97E-14	1	RPL13
2.92E-18	0.307074	0.29	0.041	9.79E-14	1	GPRC5C
3.22E-18	0.317945	0.312	0.046	1.08E-13	1	MCUB
3.24E-18	0.350339	1	0.997	1.09E-13	1	RPS16
5.47E-18	0.383844	1	0.985	1.83E-13	1	RPL23A

6.69E-18	0.317444	1	0.999	2.24E-13	1	RPL3
7.50E-18	0.385905	1	0.996	2.51E-13	1	RPL7
9.30E-17	0.534653	0.989	0.806	3.12E-12	1	GSTP1
1.23E-16	0.435451	0.43	0.097	4.13E-12	1	EMP3
1.34E-16	0.252927	0.151	0.008	4.49E-12	1	TCF7L1
2.66E-16	0.313172	1	0.997	8.92E-12	1	RPS7
4.39E-16	0.304722	1	0.999	1.47E-11	1	RPS27A
4.68E-16	0.358927	1	1	1.57E-11	1	RPS6
6.83E-16	0.46458	0.484	0.138	2.29E-11	1	TIMP1
8.85E-16	0.358071	1	0.992	2.97E-11	1	RPLP0
9.69E-16	0.321936	1	0.997	3.25E-11	1	RPS25
1.12E-15	0.741856	0.72	0.329	3.77E-11	1	MEST
1.37E-15	0.325226	1	0.997	4.60E-11	1	RPL35
1.61E-15	0.313219	1	0.997	5.39E-11	1	RPL29
2.10E-15	0.315151	1	0.997	7.05E-11	1	RPL19
2.20E-15	0.299418	0.387	0.084	7.37E-11	1	NES
3.83E-15	0.343139	1	0.996	1.29E-10	1	RPL10A
3.96E-15	0.345884	0.376	0.083	1.33E-10	1	FZD2
5.40E-15	0.315812	1	1	1.81E-10	1	RPS12
5.90E-15	0.512443	0.43	0.117	1.98E-10	1	ANXA2
1.03E-14	0.326745	0.312	0.06	3.45E-10	1	FKBP10
1.41E-14	0.477065	0.495	0.147	4.73E-10	1	PHLDA1
1.67E-14	0.495381	1	0.841	5.60E-10	1	EEF1D
3.11E-14	0.699179	0.806	0.458	1.04E-09	1	TTYH1

Ribosomal

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
4.39E-46	0.873794	0.147	0.033	1.47E-41	7	GDF15
4.57E-40	1.253904	0.532	0.377	1.53E-35	7	CDKN1A
3.54E-26	0.709275	0.572	0.459	1.19E-21	7	CKS2
1.12E-23	0.547539	0.168	0.066	3.77E-19	7	SRRT
5.95E-21	0.66991	0.423	0.308	1.99E-16	7	ADRM1
9.52E-21	0.506334	0.179	0.077	3.19E-16	7	PMAIP1
2.46E-20	0.897737	0.381	0.276	8.24E-16	7	GADD45A
6.01E-19	0.601215	0.4	0.287	2.02E-14	7	ZMAT3
1.09E-17	0.604866	0.329	0.218	3.65E-13	7	HNRNPM
4.03E-17	0.644047	0.265	0.161	1.35E-12	7	HSPH1
3.92E-11	0.328198	0.045	0.148	1.32E-06	7	CYB5D2
1.56E-08	1.061254	0.097	0.2	5.25E-04	7	SYP
3.81E-08	0.346051	0.123	0.236	1.28E-03	7	KLHL7
9.81E-03	0.253098	0.652	0.768	1.00E+00	7	ATP6V1G1

Pre-thalamic progenitors

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
1.18E-293	0.623602	0.286	0.011	3.97E-289	10	GSX2
3.67E-278	1.955124	0.75	0.115	1.23E-273	10	SFRP2
2.22E-204	0.805884	0.451	0.046	7.43E-200	10	VCAM1
4.01E-187	0.992037	0.76	0.148	1.35E-182	10	WNT7B
3.46E-184	1.59687	0.865	0.257	1.16E-179	10	CCND1
1.78E-181	0.382868	0.319	0.025	5.98E-177	10	AC092958.1
8.91E-163	1.568182	0.638	0.123	2.99E-158	10	ASCL1
9.54E-162	0.254202	0.197	0.01	3.20E-157	10	CA8
5.82E-160	0.307958	0.211	0.012	1.95E-155	10	THSD4
1.80E-155	0.264003	0.227	0.014	6.04E-151	10	PLAC9
8.78E-155	0.696607	0.618	0.108	2.95E-150	10	FEZF1-AS1
4.15E-151	0.269812	0.253	0.019	1.39E-146	10	ANOS1
1.14E-134	0.32567	0.257	0.022	3.83E-130	10	SOX21
4.11E-131	1.295155	0.993	0.883	1.38E-126	10	VIM
2.22E-130	0.591597	0.586	0.111	7.44E-126	10	FEZF1
2.78E-127	0.483601	0.365	0.047	9.33E-123	10	ACTA2
5.49E-125	0.3438	0.289	0.03	1.84E-120	10	GLIS3
9.55E-124	1.321834	0.987	0.682	3.20E-119	10	FABP7
3.56E-122	0.31647	0.312	0.035	1.20E-117	10	LINC01833
3.15E-119	1.200781	0.447	0.077	1.06E-114	10	CTGF
2.44E-118	0.567676	0.385	0.057	8.19E-114	10	FAM181A
3.27E-116	0.499021	0.375	0.053	1.10E-111	10	NPTX2
2.83E-115	0.734192	0.595	0.133	9.50E-111	10	TCF7L2
2.38E-100	0.677409	0.658	0.181	7.99E-96	10	PLAGL1
3.53E-97	0.688108	0.349	0.056	1.18E-92	10	PTX3
5.51E-95	0.736239	0.362	0.061	1.85E-90	10	AL139246.5
9.10E-92	0.49469	0.359	0.059	3.05E-87	10	ZNF385D
9.49E-91	0.338515	0.26	0.033	3.18E-86	10	TOX
1.29E-80	0.457992	0.464	0.106	4.33E-76	10	PDLIM3
6.92E-80	0.742127	0.875	0.447	2.32E-75	10	GNG5
2.96E-79	0.595065	0.602	0.171	9.92E-75	10	ARL4A
8.95E-79	0.371604	0.349	0.064	3.00E-74	10	VEPH1
1.32E-78	0.77133	0.493	0.122	4.43E-74	10	LHX5-AS1
1.34E-77	0.889392	0.931	0.476	4.51E-73	10	SOX2
2.13E-77	0.686998	0.595	0.188	7.14E-73	10	LITAF
1.28E-75	0.858055	0.845	0.43	4.28E-71	10	GPM6B

1.59E-75	0.403209	0.444	0.1	5.32E-71	10	CXXC4
2.04E-75	0.566641	0.5	0.13	6.85E-71	10	LIX1
6.67E-75	0.689821	0.622	0.186	2.24E-70	10	NTRK2
7.33E-75	0.85158	0.914	0.583	2.46E-70	10	ATXN10
1.12E-73	0.41617	0.418	0.093	3.77E-69	10	EPHB2
4.53E-71	0.597077	0.602	0.186	1.52E-66	10	MGST1
1.12E-70	0.66081	0.431	0.107	3.76E-66	10	CA2
5.89E-70	0.744137	0.812	0.399	1.98E-65	10	ATP1B3
9.05E-70	0.658601	0.977	0.558	3.04E-65	10	NR2F1
1.25E-69	1.062216	0.931	0.569	4.18E-65	10	ID1
1.44E-67	0.703142	0.678	0.259	4.82E-63	10	NES
1.16E-65	0.429231	0.355	0.077	3.88E-61	10	TCIM
3.22E-63	0.328974	0.28	0.051	1.08E-58	10	AC004540.2
3.37E-63	0.594646	0.862	0.361	1.13E-58	10	NR2F2
2.94E-62	0.746169	0.191	0.026	9.84E-58	10	TAGLN
3.01E-62	0.725143	0.914	0.637	1.01E-57	10	CNN3
1.54E-61	0.815382	0.839	0.536	5.15E-57	10	SMS
8.21E-60	1.227122	0.757	0.407	2.75E-55	10	PTN
1.39E-59	1.084479	0.868	0.495	4.66E-55	10	ID3
4.69E-59	0.718309	0.901	0.488	1.57E-54	10	PAX6
6.28E-59	0.273501	0.293	0.058	2.11E-54	10	EPHA2
2.09E-58	0.262147	0.227	0.037	7.02E-54	10	PLAT
9.03E-58	0.453711	0.214	0.034	3.03E-53	10	GDF15
2.33E-57	0.501722	0.533	0.171	7.80E-53	10	S100A13
3.73E-55	0.463958	0.605	0.217	1.25E-50	10	QKI
1.55E-53	0.378865	0.47	0.143	5.21E-49	10	TP53TG1
3.37E-53	0.551122	0.977	0.865	1.13E-48	10	GSTP1
9.24E-53	0.265325	0.312	0.07	3.10E-48	10	BCAT1
1.07E-52	0.799665	0.349	0.092	3.60E-48	10	CYR61
1.48E-52	0.344226	0.388	0.103	4.96E-48	10	ZIC1
3.01E-52	0.617083	0.845	0.474	1.01E-47	10	CLIC1
5.39E-51	0.622492	0.921	0.681	1.81E-46	10	MGST3
1.30E-50	0.586589	0.576	0.219	4.36E-46	10	SOX3
1.70E-50	0.519294	0.342	0.089	5.71E-46	10	GATM
4.24E-50	0.413815	0.414	0.12	1.42E-45	10	TPPP3
4.36E-50	0.285259	0.266	0.056	1.46E-45	10	NR2E1
1.47E-49	0.397771	0.322	0.079	4.94E-45	10	WNT7A
4.56E-49	0.615445	0.609	0.259	1.53E-44	10	MYL12A
4.15E-48	0.582911	0.684	0.292	1.39E-43	10	NKAIN4

4.16E-47	0.379827	0.352	0.095	1.39E-42	10	CHIC2
2.33E-45	0.369737	0.5	0.171	7.83E-41	10	METRN
7.33E-45	0.442705	0.678	0.292	2.46E-40	10	TMEM98
9.72E-45	0.500167	0.632	0.264	3.26E-40	10	CRB2
1.44E-44	0.427659	0.309	0.08	4.82E-40	10	PURPL
3.02E-44	0.443539	0.569	0.22	1.01E-39	10	RAB13
6.20E-43	0.410676	0.395	0.127	2.08E-38	10	TUBB6
8.22E-42	0.285263	0.319	0.087	2.76E-37	10	SLC9A3R1
1.46E-41	0.388436	0.477	0.171	4.89E-37	10	C12orf75
1.56E-41	0.497568	0.977	0.765	5.24E-37	10	RPS27L
7.27E-40	0.480163	0.786	0.45	2.44E-35	10	SAT1
1.13E-39	0.444748	0.859	0.534	3.80E-35	10	SRI
1.68E-39	0.619205	0.638	0.307	5.62E-35	10	FGFBP3
1.76E-39	0.304115	0.428	0.14	5.90E-35	10	ADGRV1
7.19E-39	0.40718	0.484	0.18	2.41E-34	10	GPC3
1.50E-38	0.576803	0.918	0.728	5.04E-34	10	FDPS
3.77E-38	0.637049	0.711	0.376	1.26E-33	10	CDKN1A
1.04E-37	0.250823	0.286	0.076	3.48E-33	10	NR2F2-AS1

Cajal-Retzius neurons

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fctr>	<chr>
0.00E+00	2.40032	0.991	0.041	0.00E+00	11	LHX1
0.00E+00	2.062884	0.862	0.047	0.00E+00	11	MAB21L1
0.00E+00	1.782932	0.798	0.033	0.00E+00	11	LHX1-DT
0.00E+00	0.895627	0.339	0.002	0.00E+00	11	TP73
1.74E-177	0.930623	0.404	0.017	5.85E-173	11	KCNC2
1.26E-172	1.257878	0.615	0.043	4.21E-168	11	LHX5
5.04E-139	1.142936	0.596	0.053	1.69E-134	11	PGF
2.22E-119	0.478565	0.22	0.007	7.43E-115	11	EBF3
1.37E-101	1.43541	0.789	0.125	4.59E-97	11	LHX5-AS1
1.85E-95	1.227312	0.56	0.067	6.22E-91	11	GABRA2
1.65E-93	1.084545	0.431	0.039	5.55E-89	11	RELN
3.16E-89	0.522682	0.202	0.009	1.06E-84	11	VSTM2A
2.11E-87	1.355239	0.679	0.108	7.06E-83	11	NHLH2
3.23E-75	0.605772	0.303	0.023	1.08E-70	11	SAMD3
1.10E-74	0.366778	0.174	0.008	3.70E-70	11	SYNDIG1
1.11E-68	1.084027	0.578	0.098	3.73E-64	11	PIDD1
1.67E-53	0.567786	0.33	0.039	5.61E-49	11	GALNT8
3.30E-50	0.963671	0.642	0.152	1.11E-45	11	ELAVL2
8.40E-49	1.086745	0.532	0.107	2.82E-44	11	NEUROD1
1.17E-39	0.380179	0.147	0.011	3.91E-35	11	CDH13
3.92E-39	1.250627	0.936	0.755	1.31E-34	11	TXN
1.11E-33	0.404661	0.183	0.019	3.71E-29	11	PCP4L1
1.16E-32	1.027618	0.826	0.369	3.90E-28	11	NR2F2
2.92E-32	0.824976	0.936	0.676	9.81E-28	11	ETFB
4.63E-32	0.768347	1	0.8	1.55E-27	11	MLLT11
2.93E-31	0.339249	0.147	0.013	9.81E-27	11	CNTNAP5
8.59E-30	0.697967	0.284	0.049	2.88E-25	11	SHISA2
2.06E-28	0.84392	0.936	0.578	6.91E-24	11	EMX2
1.27E-27	0.31081	0.156	0.017	4.25E-23	11	VSNL1
2.17E-27	0.661938	0.312	0.061	7.28E-23	11	EBF2
5.60E-27	0.819411	0.817	0.434	1.88E-22	11	KIF5C
9.46E-26	0.667163	0.468	0.127	3.17E-21	11	NHLH1
2.07E-25	0.698134	0.963	0.53	6.96E-21	11	RTN1
2.92E-25	0.773822	0.578	0.202	9.78E-21	11	NRXN1
4.76E-25	0.900597	0.532	0.187	1.60E-20	11	ZNF503
6.13E-25	0.486559	0.165	0.021	2.06E-20	11	CPNE4

6.30E-25	0.794293	0.771	0.38	2.11E-20	11	ATP6V1G2
3.04E-24	0.689699	0.55	0.193	1.02E-19	11	CDK5R1
3.55E-24	0.78112	0.725	0.312	1.19E-19	11	NRN1
5.61E-24	0.675761	1	0.797	1.88E-19	11	TUBB2A
1.24E-23	0.558309	0.991	0.871	4.17E-19	11	UCHL1
1.94E-23	0.538859	0.275	0.056	6.51E-19	11	CNR1
2.35E-23	0.316919	0.165	0.022	7.88E-19	11	FNDC5
3.44E-23	0.426533	0.239	0.043	1.15E-18	11	DLG2
9.87E-23	0.656423	0.963	0.753	3.31E-18	11	TERF2IP
1.03E-21	0.607357	0.917	0.698	3.44E-17	11	NREP
6.61E-20	0.617852	0.908	0.591	2.22E-15	11	CNTNAP2
2.22E-19	0.706072	0.532	0.207	7.43E-15	11	PPP3CA
5.36E-19	0.540768	0.991	0.615	1.80E-14	11	STMN2
8.89E-19	0.640648	0.688	0.323	2.98E-14	11	CDKN2D
9.69E-19	0.434409	0.165	0.026	3.25E-14	11	SPINK5
1.09E-18	0.623741	0.872	0.591	3.67E-14	11	CRMP1
1.38E-18	0.846463	0.495	0.192	4.62E-14	11	PCDH9
1.65E-18	0.510254	0.312	0.083	5.53E-14	11	CHST8
3.62E-18	0.706137	0.404	0.141	1.21E-13	11	AATF
5.86E-18	0.588376	0.862	0.503	1.96E-13	11	RAB3A
7.06E-18	0.612031	0.523	0.199	2.37E-13	11	TBR1
8.32E-18	0.602637	0.725	0.357	2.79E-13	11	OLFM1
8.60E-18	0.42818	0.202	0.04	2.88E-13	11	RGS7
1.77E-17	0.618023	0.716	0.404	5.94E-13	11	CADM1
2.50E-17	0.682572	0.798	0.527	8.38E-13	11	NOVA1
3.78E-17	0.4845	0.284	0.074	1.27E-12	11	SLC17A6
4.53E-17	0.343866	0.147	0.023	1.52E-12	11	LHX9
5.79E-17	0.664969	0.642	0.309	1.94E-12	11	INA
8.47E-17	0.628776	0.706	0.384	2.84E-12	11	ELAVL3
1.19E-16	0.492946	0.56	0.225	3.98E-12	11	SNAP25
3.86E-16	0.575859	0.459	0.179	1.29E-11	11	OPTN
5.21E-16	0.72897	0.55	0.262	1.75E-11	11	IER3
5.35E-16	0.514782	0.312	0.092	1.79E-11	11	MYT1L
5.92E-16	0.665897	0.505	0.21	1.99E-11	11	ENC1
6.22E-16	0.457125	0.257	0.066	2.08E-11	11	LRRC26
1.22E-15	0.514944	0.853	0.631	4.09E-11	11	GDI1
6.84E-15	0.512303	0.697	0.34	2.29E-10	11	NSG2
8.18E-15	0.399894	0.275	0.077	2.74E-10	11	GABRG2
9.89E-15	0.464498	0.972	0.824	3.32E-10	11	RTN4

1.15E-14	0.465248	0.835	0.519	3.86E-10	11	TAGLN3
2.06E-14	0.400218	0.239	0.061	6.91E-10	11	TMEFF2
2.41E-14	0.473442	0.936	0.725	8.08E-10	11	GPM6A
4.21E-14	0.367129	0.257	0.07	1.41E-09	11	EDIL3
4.69E-14	0.625526	0.349	0.12	1.57E-09	11	FEZF1
8.40E-14	0.517155	0.394	0.152	2.82E-09	11	RGMB
9.38E-14	0.474693	0.339	0.123	3.15E-09	11	EFNB2
9.55E-14	0.443872	0.945	0.8	3.20E-09	11	DDAH2
4.40E-13	0.440117	0.514	0.221	1.47E-08	11	SCG3
4.44E-13	0.712586	0.78	0.561	1.49E-08	11	JUN
4.49E-13	0.426833	0.22	0.06	1.51E-08	11	SLITRK4
5.69E-13	0.453901	0.771	0.489	1.91E-08	11	NSG1
8.84E-13	0.685568	0.917	0.786	2.96E-08	11	BTG1
1.17E-12	0.813157	0.477	0.234	3.92E-08	11	RGS16
1.84E-12	0.311521	0.193	0.048	6.19E-08	11	TMEM151B
2.37E-12	0.421165	0.936	0.801	7.94E-08	11	FXYD6
2.39E-12	0.568872	0.771	0.527	8.03E-08	11	BLCAP
2.79E-12	0.574796	0.404	0.178	9.35E-08	11	ING2

Appendix 5: scRNA-Seq differentially expressed genes (by cluster)

Cortical Progenitors

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
DLK1	4.7E-37	-0.37308	0.343	0.558	9.16E-33
CRABP1	2.19E-45	-0.35739	0.474	0.693	4.27E-41
ID1	1.99E-25	-0.28057	0.891	0.919	3.87E-21
SOX4	7.46E-18	-0.22413	0.761	0.846	1.45E-13
COL9A3	2.51E-27	-0.22288	0.361	0.548	4.88E-23
CRABP2	1.69E-20	-0.22215	0.672	0.772	3.29E-16
CXCR4	1.52E-20	-0.22049	0.372	0.53	2.96E-16
ID2	2.74E-15	-0.21778	0.75	0.817	5.34E-11
SFRP1	9.41E-21	0.206754	0.801	0.692	1.83E-16

Immature cortical neurons

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
CALB2	5.46E-11	-0.39431	0.229	0.432	1.06E-06
NRN1	2.14E-10	-0.32169	0.277	0.478	4.16E-06
LHX2	1.51E-12	-0.30417	0.811	0.928	2.94E-08
CRABP1	4.47E-09	-0.22044	0.971	0.986	8.72E-05
THSD7A	5.59E-07	-0.20591	0.191	0.353	0.010893
DOK5	1.06E-07	0.264979	0.576	0.399	0.002068
SPINT2	3.26E-06	-0.20628	0.539	0.665	0.063453
TMEM97	9.89E-06	0.201551	0.701	0.572	0.192705
MEST	1.27E-05	-0.20454	0.304	0.45	0.246653
HMGCS1	8.84E-05	0.220861	0.836	0.784	1
DLK1	0.000215	0.232512	0.507	0.378	1
CRABP2	0.001298	-0.20883	0.42	0.511	1

Intermediate cortical neurons

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
CALB2	1.05E-10	-0.4569	0.413	0.689	2.04E-06
NRN1	1.86E-06	-0.37601	0.523	0.682	0.036171
LHX2	2.42E-06	-0.30698	0.656	0.788	0.047175
RPL32	4.99E-10	0.200688	1	1	9.72E-06
RPL34	1.87E-08	0.202654	1	1	0.000364
RPL39	7.79E-09	0.224143	1	1	0.000152
NEUROD6	3.23E-06	0.224673	0.987	0.947	0.062909
ID1	5.05E-06	-0.30757	0.508	0.689	0.098514
HMGCS1	5.61E-06	0.250352	0.906	0.795	0.1093
SH3BGRL3	6.66E-06	-0.26715	0.561	0.735	0.129822
RTN1	1.26E-05	-0.21214	0.967	0.987	0.245005
CNTNAP2	1.71E-05	0.213623	0.885	0.689	0.333194
RASD1	2.88E-05	-0.32972	0.329	0.49	0.561907
CRABP1	3.98E-05	-0.26368	0.809	0.914	0.776468
LAPTM4A	4.16E-05	-0.23614	0.64	0.788	0.810096
CELF4	8.09E-05	-0.21879	0.735	0.821	1
MSMO1	8.24E-05	0.21401	0.941	0.854	1
TMEM97	0.000145	0.203356	0.577	0.384	1
CSRP2	0.000227	0.293173	0.571	0.43	1
ARL2	0.000379	-0.23006	0.372	0.503	1
NOVA1	0.00109	-0.21337	0.398	0.503	1
HPCAL1	0.004942	-0.20798	0.439	0.536	1

Mature cortical neurons

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
HMGCS1	2.4E-44	0.634422	0.902	0.655	4.68E-40
LMO3	1E-15	0.633618	0.439	0.218	1.95E-11
CSRP2	1.44E-36	0.553706	0.93	0.762	2.81E-32
ACAT2	2.77E-28	0.461404	0.847	0.605	5.39E-24
MSMO1	5.7E-29	0.450866	0.935	0.782	1.11E-24
FDFT1	4.33E-34	0.429782	0.976	0.878	8.43E-30
FDPS	2.91E-23	0.424192	0.872	0.722	5.68E-19
NEUROD2	2.03E-22	0.365703	0.967	0.859	3.95E-18
SQLE	7.6E-18	0.326197	0.875	0.722	1.48E-13
IDI1	5.02E-18	0.29879	0.904	0.806	9.77E-14
TMEM97	5.06E-12	0.276397	0.568	0.387	9.86E-08
CNTNAP2	2.79E-16	0.26541	0.993	0.96	5.44E-12
MVD	4.65E-14	0.261297	0.62	0.404	9.06E-10
INSIG1	2.27E-12	0.240814	0.665	0.444	4.42E-08
HMGCR	8.13E-12	0.232921	0.604	0.392	1.58E-07
DHCR7	6.06E-12	0.231039	0.545	0.325	1.18E-07
TSC22D1	2.48E-10	0.226335	0.779	0.64	4.83E-06
MAGEH1	1.15E-12	-0.20261	0.372	0.596	2.24E-08
NHLH2	9.36E-10	-0.20356	0.315	0.486	1.82E-05
ARID5B	1.29E-11	-0.20373	0.273	0.467	2.5E-07
ATP1B1	4.73E-08	-0.21097	0.515	0.655	0.000922
SLC22A17	3.99E-11	-0.21248	0.691	0.819	7.77E-07
WLS	6.71E-13	-0.21668	0.206	0.402	1.31E-08
RGS17	1.31E-08	-0.22113	0.521	0.643	0.000256
THSD7A	2.7E-08	-0.22297	0.763	0.871	0.000526
LY6H	6.33E-14	-0.22969	0.772	0.896	1.23E-09
ZIC2	7.03E-09	-0.23164	0.448	0.596	0.000137
NFIB	1.49E-09	-0.23229	0.951	0.97	2.91E-05
CELF4	7.15E-15	-0.25269	0.838	0.926	1.39E-10
SYNPR	4.04E-17	-0.27523	0.165	0.39	7.88E-13
NRN1	8.86E-11	-0.28032	0.73	0.878	1.73E-06
LHX2	5.21E-08	-0.28068	0.545	0.653	0.001016
IGFBP5	2.42E-06	-0.31858	0.614	0.73	0.047202
DIRAS3	6.62E-09	-0.35329	0.301	0.454	0.000129
BHLHE22	5.08E-24	-0.40112	0.283	0.576	9.9E-20
NFIA	9.44E-29	-0.40393	0.865	0.943	1.84E-24
LAPTM4A	4.44E-30	-0.40885	0.586	0.834	8.65E-26
CRABP1	1.29E-23	-0.44175	0.49	0.734	2.52E-19

NTS	6.79E-22	-0.59422	0.291	0.571	1.32E-17
CALB2	5.74E-51	-0.74772	0.668	0.933	1.12E-46

Cortical Hem

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
C1orf61	4.83E-16	-0.30728	0.549	0.711	9.41E-12
PKM	6.5E-21	-0.25957	0.887	0.949	1.27E-16
NELL2	5.6E-11	-0.22537	0.524	0.637	1.09E-06
WSB1	1.41E-12	-0.21865	0.782	0.868	2.74E-08
NMU	8.47E-11	0.23568	0.749	0.633	1.65E-06
DLK1	9.44E-07	0.522396	0.507	0.41	0.018403
NMB	6.99E-06	0.231784	0.427	0.348	0.13629
SPARCL1	0.000559	0.354719	0.369	0.297	1
TRH	0.784945	0.214858	0.52	0.552	1

Choroid Plexus

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
PKM	1.2E-12	-0.29744	0.82	0.925	2.35E-08
C1orf61	2.37E-08	-0.2969	0.356	0.523	0.000462
OLFM3	1.73E-11	-0.23851	0.243	0.448	3.37E-07
CNTNAP2	2.42E-07	-0.22437	0.804	0.853	0.004719
TUBA1A	1.97E-13	-0.22239	0.997	1	3.84E-09
HNRNPC	9.28E-09	-0.2024	0.873	0.934	0.000181
TRPM3	1.26E-09	0.346392	0.604	0.42	2.47E-05
DLK1	5.28E-13	0.683516	0.774	0.624	1.03E-08
MASP1	4.43E-06	-0.20173	0.318	0.463	0.086264
RSPO3	9.84E-06	-0.23752	0.492	0.601	0.191719
PGM5P4-AS1	1.22E-05	0.275857	0.421	0.287	0.237381
HSPA5	1.25E-05	-0.22825	0.609	0.718	0.243999
AL596223.2	1.82E-05	-0.24097	0.343	0.466	0.355136
CLDN5	8.78E-05	-0.28853	0.683	0.741	1
NMB	0.000218	0.2932	0.698	0.632	1
TTR	0.000477	0.204819	1	1	1

Pre-thalamus

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
NEFM	1.85E-13	0.225687	0.261	0.144	3.61E-09

Pre-thalamic eminence (EMX1 sub-cluster)

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
CRABP1	1.29E-10	-0.49347	0.511	0.804	2.51E-06
RPL39	2.79E-06	0.225527	1	1	0.054403
ERG28	1.51E-05	0.248921	0.534	0.307	0.293827
SQLE	2.83E-05	0.264131	0.722	0.503	0.551288
HMGCS1	5.38E-06	0.317529	0.767	0.558	0.104832
EIF4A2	0.000112	0.217506	0.983	0.92	1
FDPS	0.000132	0.256079	0.722	0.54	1
MAGEH1	0.000253	-0.2225	0.432	0.613	1
LMO3	0.000417	0.376933	0.301	0.153	1
CALB2	0.000645	-0.27776	0.233	0.405	1
YWHAB	0.000769	-0.21242	0.648	0.748	1
FDFT1	0.000946	0.244242	0.886	0.859	1
TMEM161B-AS1	0.001162	0.229131	0.682	0.521	1
HMGCR	0.001206	0.22819	0.432	0.264	1
RABAC1	0.001906	-0.20099	0.46	0.62	1
RGS17	0.002042	-0.21585	0.58	0.699	1
MSMO1	0.002483	0.260594	0.83	0.761	1
INSIG1	0.004429	0.219532	0.5	0.356	1
NTS	0.008723	-0.32526	0.403	0.534	1
PAX6	0.012766	0.30942	0.369	0.252	1
NTM	0.024911	-0.20427	0.369	0.466	1
SIRT2	0.135492	0.268713	0.295	0.227	1

Pre-thalamic eminence (LHX5 sub-cluster)

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
CRABP1	4.27E-17	-0.54661	0.438	0.666	8.33E-13
LHX5-AS1	4.44E-22	-0.42972	0.356	0.668	8.65E-18
NOVA1	1.83E-09	-0.26911	0.632	0.779	3.56E-05
RSPO3	6.58E-09	-0.22509	0.091	0.219	0.000128
MSMO1	7.95E-08	0.221041	0.856	0.8	0.001549
IDI1	9.66E-10	0.255918	0.819	0.74	1.88E-05
FDFT1	6.55E-16	0.307114	0.931	0.87	1.28E-11
SQLE	1.08E-11	0.320354	0.683	0.527	2.1E-07
HMGCS1	4.89E-12	0.333254	0.751	0.64	9.54E-08
PTN	5.55E-05	-0.23588	0.39	0.505	1
LMO4	0.00067	0.264306	0.375	0.293	1
PBX3	0.000705	0.212271	0.347	0.258	1
NEFM	0.00104	-0.23909	0.345	0.425	1
CTNNB1	0.001598	-0.23158	0.327	0.41	1

Pre-thalamic progenitors

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
CALB1	2.46E-07	-0.60841	0.058	0.279	0.004793
CRISPLD1	4.01E-06	-0.29938	0.05	0.242	0.07807
MYL12A	4.47E-06	-0.42754	0.475	0.661	0.08705
PRDX2	1.3E-05	-0.20231	0.978	0.982	0.253589
GLIS3	1.81E-05	-0.20093	0.144	0.364	0.352262
ANXA1	4.71E-05	-0.34906	0.094	0.279	0.917133
PLAT	5.21E-05	-0.22563	0.108	0.297	1
PTN	5.24E-05	-0.41332	0.597	0.836	1
REC8	5.48E-05	-0.20077	0.094	0.273	1
GSTP1	7.45E-05	-0.24799	0.986	1	1
TUBB6	7.54E-05	-0.25906	0.245	0.461	1
MGST3	7.84E-05	-0.22679	0.871	0.939	1
CTGF	0.0001	-0.58565	0.317	0.527	1
WLS	0.000104	-0.30171	0.504	0.697	1
ANXA2	0.000108	-0.38668	0.388	0.588	1
ITM2C	0.000139	-0.23399	0.309	0.527	1
TPM1	0.000193	-0.22567	0.223	0.418	1
MYL6	0.000197	-0.2389	0.986	0.982	1

TCIM	0.000213	-0.23399	0.23	0.418	1
SNRPB2	0.000246	-0.20594	0.518	0.739	1

Ribosomal cluster

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
RPLP1	1.98E-09	0.283549	1	1	3.86E-05
RPS10	4.06E-09	0.260849	1	1	7.92E-05
RPS12	1.87E-08	0.276056	1	1	0.000364
RPS6	4.75E-08	0.294619	1	0.997	0.000925
RPL22	6.38E-08	0.245723	1	0.997	0.001244
RPL41	7.26E-08	0.237062	1	1	0.001414
RPL18A	7.56E-08	0.201834	1	1	0.001473
MAGEH1	7.58E-08	-0.27582	0.094	0.274	0.001477
RPS27A	8.89E-08	0.223479	1	1	0.001732
RPL39	1.12E-07	0.276173	1	1	0.002192
RPS24	1.66E-07	0.293632	1	1	0.003239
RPL34	1.68E-07	0.291534	1	1	0.003282
RPL13	1.82E-07	0.238352	1	1	0.00354
RPS17	1.85E-07	0.229503	1	0.997	0.003603
RPS8	1.95E-07	0.265257	1	1	0.003807
RPL11	2.17E-07	0.220271	1	1	0.004228
RPS15A	2.81E-07	0.250226	1	1	0.005471
RPL37A	2.96E-07	0.219047	1	1	0.005763
RPS3A	3.52E-07	0.240399	1	1	0.006851
RPL37	4.16E-07	0.231342	1	1	0.008106
RPS23	6.34E-07	0.245991	1	1	0.012356
RPL31	6.53E-07	0.248929	1	1	0.012718
RPL12	1.72E-06	0.2198	1	0.997	0.033468
RPL28	1.81E-06	0.212402	1	1	0.035223
RPL36A	3.91E-06	0.24447	0.991	0.994	0.076297
RPL27	8.93E-06	0.210294	1	1	0.174103
NAP1L1	1.26E-05	0.213925	0.996	0.989	0.246275
VAMP2	1.68E-05	-0.32225	0.664	0.751	0.327854
EIF3E	2.49E-05	0.200029	0.96	0.946	0.486254
RPL30	2.97E-05	0.211855	1	1	0.578881
H3F3B	3.01E-05	-0.27448	0.996	0.994	0.586175
SRSF5	3.92E-05	-0.22681	0.457	0.619	0.764644

Cajal-Retzius neurons

Gene	p_val	avg_logFC	% WT cells expressing the gene	% KO cells expressing the gene	p_val_adj
P3H4	0.000345	0.335094	0.455	0.126	1
UNC5B	0.000494	0.218434	0.227	0.023	1
NQO2	0.000928	0.353214	0.455	0.149	1
PI4KB	0.001199	0.243464	0.318	0.069	1
TMEM259	0.001254	0.285823	0.5	0.172	1
CEP170	0.001287	0.345246	0.591	0.253	1
UBA52	0.001471	0.255634	1	0.989	1
DCTN3	0.001541	0.402134	0.864	0.506	1
MRPL54	0.001616	0.239886	0.591	0.218	1
MT-ATP6	0.001839	0.37124	0.955	0.977	1
TXNDC9	0.0019	0.276253	0.409	0.126	1
ARFRP1	0.00234	0.254274	0.273	0.057	1
FAM174A	0.00239	0.244505	0.364	0.103	1
ATP6V1G1	0.002519	0.330113	1	0.816	1
NDUFS4	0.002576	0.329958	0.773	0.414	1
CYBC1	0.002627	0.220373	0.273	0.057	1
CNPY3	0.002796	0.226982	0.5	0.184	1
ZNF385A	0.003039	0.209562	0.273	0.057	1
FAM204A	0.003123	0.357479	0.682	0.356	1
COX4I1	0.003153	0.240615	1	1	1
CNIH4	0.00319	0.2464	0.409	0.126	1
WDR54	0.003287	0.411428	0.773	0.483	1
FAM53C	0.003315	0.343264	0.409	0.138	1
SVBP	0.003483	0.341851	0.818	0.609	1
TALDO1	0.003627	0.323881	0.591	0.276	1
SND1	0.003741	0.276253	0.5	0.195	1
COA4	0.003979	0.265891	0.318	0.08	1
HSP90B1	0.004325	0.325044	0.682	0.368	1
SNRPF	0.004439	0.263514	0.864	0.46	1
AC090204.1	0.004453	0.234142	0.364	0.115	1
NDUFB10	0.00458	0.352695	0.864	0.759	1
SLC25A6	0.005099	0.387479	0.955	0.943	1

Appendix 6: HARE5 CADD scores